On the Distribution of Spacings Between Zeros of the Zeta Function

A. M. Odlyzko

On the Distribution of Spacings Between Zeros of the Zeta Function

By A. M. Odlyzko

Dedicated to Daniel Shanks on the occasion of his seventieth birthday

Abstract. A numerical study of the distribution of spacings between zeros of the Riemann zeta function is presented. It is based on values for the first 10^5 zeros and for zeros number $10^{12} + 1$ to $10^{12} + 10^5$ that are accurate to within $\pm 10^{-8}$, and which were calculated on the Cray-1 and Cray X-MP computers. This study tests the Montgomery pair correlation conjecture as well as some further conjectures that predict that the zeros of the zeta function behave like eigenvalues of random Hermitian matrices. Matrices of this type are used in modeling energy levels in physics, and many statistical properties of their eigenvalues are known. The agreement between actual statistics for zeros of the zeta function and conjectured results is generally good, and improves at larger heights. Several initially unexpected phenomena were found in the data and some were explained by relating them to the primes.

1. Introduction. The Riemann Hypothesis (RH) has been of central interest to number theorists for a long time, and many unsuccessful attempts have been made to either prove or disprove it by analytic methods. A series of large computations have also been made, culminating in the recent verification [52] that the RH holds for the first $1.5 \cdot 10^9$ zeros, an effort that involved over a thousand hours on a modern supercomputer.

A proof of the RH would lead to tremendous improvements in estimates for various arithmetic functions, such as the difference between $\pi(x)$, the number of primes $\leqslant x$, and $\mathrm{Li}(x)$, the logarithmic integral of x. However, many other questions would remain open, such as the precise magnitude of the largest gap between consecutive primes below a given bound. Answers to such questions depend on a much more detailed knowledge of the distribution of zeros of the zeta function than is given by the RH. Relatively little work has been devoted to the precise distribution of the zeros. The main reason for the lack of research in this area was undoubtedly the feeling that there was little to be gained from studying problems harder than the RH if the RH itself could not be proved.

A major step towards a detailed study of the distribution of zeros of the zeta function was made by H. L. Montgomery [56], [57]. Under the assumption of the

RH, he showed that if we define

$$(1.1) \qquad F(\alpha, T) = 2\pi (T \log T)^{-1} \sum_{\substack{0 < \gamma \leqslant T \\ 0 < \gamma' \leqslant T}} T^{i\alpha(\gamma - \gamma')} \frac{4}{4 + (\gamma - \gamma')^2}$$

for α and T real, $T \geqslant 2$, where $\frac{1}{2} + i\gamma$ and $\frac{1}{2} + i\gamma'$ denote nontrivial zeros of the zeta function, then

$$(1.2) \qquad F(\alpha, T) = (1 + o(1))T^{-2\alpha}\log T + \alpha + o(1) \quad \text{as } T \to \infty,$$

uniformly for $0 \leqslant \alpha \leqslant 1$. Montgomery also presented heuristic arguments which suggested that

$$(1.3) \qquad F(\alpha, T) = 1 + o(1) \quad \text{as } T \to \infty$$

uniformly for $\alpha \in [a, b]$, where $1 \leqslant a < b < \infty$ are any constants. If the conjecture (1.3) were true, then the following estimate, known as the Montgomery pair correlation conjecture [56] would follow:

CONJECTURE. *For fixed* $0 < \alpha < \beta < \infty$,

$$(1.4) \qquad \frac{\left|\left\{(\gamma, \gamma'): 0 < \gamma, \gamma' \leqslant T, 2\pi\alpha(\log T)^{-1} \leqslant \gamma - \gamma' \leqslant 2\pi\beta(\log T)^{-1}\right\}\right|}{T(\log T)/(2\pi)}$$

$$\sim \int_{\alpha}^{\beta} \left(1 - \left(\frac{\sin \pi u}{\pi u}\right)^2\right) du$$

as $T \to \infty$.

The above conjecture is very striking, particularly in predicting that small gaps between zeros of the zeta function occur very infrequently. This behavior is very different from that of many other number-theoretic functions. The number of primes in short intervals, for example, is observed experimentally and is conjectured theoretically to be distributed like a Poisson random variable [29].

The conjecture (1.4) suggests some further topics for investigation. In the language of mathematical physics, this conjecture says that $1 - ((\sin \pi u)/(\pi u))^2$ is the pair correlation function of zeros of the zeta function. F. J. Dyson pointed out [56] that the Gaussian unitary ensemble (GUE) has the same pair correlation function. The GUE has been studied very extensively in mathematical physics, together with the Gaussian orthogonal ensemble (GOE) and the Gaussian symplectic ensemble (GSE), as models for distribution of energy levels in many-particle systems. The GUE consists of $n \times n$ complex Hermitian matrices of the form $A = (a_{jk})$, where $a_{jj} = 2^{1/2}\sigma_{j,j}$, $a_{jk} = \sigma_{j,k} + i\eta_{j,k}$ for $j < k$, and $a_{j,k} = \bar{a}_{k,j} = \sigma_{k,j} - i\eta_{k,j}$ for $j > k$, where the $\sigma_{j,k}$ and $\eta_{j,k}$ are independent standard normal variables. When $n \to \infty$ and the eigenvalues of the matrices of the GUE are suitably normalized, their pair correlation becomes $1 - ((\sin \pi u)/(\pi u))^2$. (The pair correlations of the GOE and GSE are different.) For more information about the GUE and the other ensembles, the reader is referred to [2], [4], [11], [41], [54], [59], [63].

The possible connection between zeros of the zeta function and eigenvalues of random matrices is of interest in number theory because of the Hilbert and Pólya conjectures [56], [59] which say that the zeros of the zeta function correspond to eigenvalues of a positive linear operator. If true, the Hilbert and Pólya conjectures would imply the RH, and some people feel that the most promising way to prove the

RH is by finding the right operator and establishing the necessary results about it. One can argue heuristically that if such an operator exists, its eigenvalues might behave like those of a random operator, which in turn might behave like the limit of a sequence of random matrices, and that therefore the zeros of the zeta function might be distributed like the eigenvalues of a large random matrix. For more on these very vague conjectures, see [3], [59]. Since Montgomery's result (1.1) is consistent with the GUE distribution of eigenvalues, one could then argue that the operator associated with the zeta function ought to be complex Hermitian, and that GUE properties of eigenvalues might apply at least approximately to the zeros of the zeta function.

Since Montgomery's proof of (1.2) (which depends on the RH) was published, several other results have been obtained. Ozluk [62] has considered a q-analogue of Montgomery's method and showed, roughly speaking, that if one considers a function similar to the $F(\alpha, T)$ of (1.1), but where one sums over zeros of many Dirichlet L-functions, then the analogue of Montgomery's conjecture (1.3) is true for $1 \leqslant \alpha \leqslant 2$. For other results about Montgomery's pair correlation conjecture and its consequences, see [30], [31], [34], [35], [36], [42], [44], [59]. Overall, though, there is very little solid theoretical evidence in favor of even the Montgomery pair correlation conjecture (1.4), and essentially none for the speculative conjectures discussed above that link zeros of the zeta function to eigenvalues of random Hermitian matrices.

One feature of the speculative arguments mentioned above is that they propose a connection between zeros of the zeta function, about which relatively little is known theoretically, and the GUE, which has been investigated very extensively and abounds in specific predictions. This makes it possible to compare data for the zeros of the zeta function against the theoretical predictions of the GUE. The purpose of this paper is to report the results of such empirical tests of the pair correlation conjecture and other GUE predictions.

The agreement between the predictions of the GUE theory and actual computed behavior of the zeros of the zeta function is quite good, especially when one takes account of the fact that the zeta function on the critical line approaches its asymptotic behavior very slowly, as is explained in Section 2. However, there are also several features of the data for the zeros that are not predicted by the GUE theory. One is that there are long-range correlations between spacings of consecutive zeros. These can be explained in terms of the primes through the use of the "explicit formulas" of prime number theory. Another is that small spacings between zeros are somewhat more common among high zeros than is expected. The deviation between the predicted and observed numbers is not large, but it is surprising, since it shows up quite early. It would be very desirable to carry out further computations at much greater heights to see if this phenomenon persists there.

This is the first study that computed large numbers of high zeros of the zeta function to substantial accuracy. The large-scale efforts to verify the RH numerically were designed only to prove that the zeros under investigation lie on the critical line, and did not produce accurate values for their positions. Some very accurate values of · zeros have also been computed in order to use them in disproving a variety of number-theoretic conjectures, but they were limited to a small number of initial

zeros (such as the roughly 100 decimal digit values for the first 2000 zeros of [60]). For a numerical investigation of the Montgomery pair correlation conjecture and the other conjectures alluded to above, it was desirable to obtain a larger sample of zeros, but only to medium accuracy. Furthermore, since there are many properties of the zeta function that are true only asymptotically, and the asymptotic behavior is often approached very slowly, it seemed desirable to compute a sample of zeros very high up. The computations on which this paper are based determined the values of the first 10^5 zeros and also of zeros number n, $10^{12} + 1 \leqslant n \leqslant 10^{12} + 10^5$, each to within $\pm 10^{-8}$, using the Cray-1 and Cray X-MP computers, respectively. In the second case, it was also verified that those zeros satisfy the RH. Zero number 10^{12} is at height approximately $2.7 \cdot 10^{11}$, and given the available computational resources and the method used, it was about as high as one could go. (A more efficient algorithm for evaluating the zeta function was invented recently [61], but it is quite complicated and has not yet been implemented.) The computations used on the order of 20 hours on the Cray X-MP, and the capabilities of this modern supercomputer were essential for the project. Several other samples of 10^5 zeros each at large heights had been computed earlier on a Cray-1, but due to a defect in the manufacturer's software, described at the end of Section 3, some of their accuracy was lost, and so they are used for only some of the statistical analyses.

The comparison of the empirical data for the zeros of the zeta function to the results that are proved for the GUE is made in Section 2. Section 3 is devoted to a description of the computation of the zeros of the zeta function and the associated error analysis. Section 4 gives some statistical information about the zeros that were found, as well as of some other zeros. This information is not directly relevant for the purposes of Section 2, but is of interest because of the comparison with the earlier computations of [10], [52]. Section 5 reports on several alternative techniques for computing the zeros that might be of use for other computations. Finally, Section 6 describes the computation of the GUE predictions.

2. Zeros and Eigenvalues. Before presenting the results of the empirical study of the zeros, we first recall some standard notation, and then discuss some known results about the zeta function and how they might affect our interpretation of the data. As usual, we consider only zeros ρ of the zeta function with $\text{Im}(\rho) > 0$, and we number them ρ_1, $\rho_2, \ldots$ (counting each according to its multiplicity) so that $0 < \text{Im}(\rho_1) \leqslant \text{Im}(\rho_2) \leqslant \ldots$. All the zeros ρ_n that have been computed so far are simple and lie on the critical line, so we can write them as $\rho_n = \frac{1}{2} + i\gamma_n$, $\gamma_n \in \mathbf{R}^+$. We have $\gamma_1 = 14.134\ldots$, $\gamma_2 = 21.022\ldots$, etc.

We let

(2.1) $\theta(t) = \arg\left[\pi^{-it/2}\Gamma\left(\tfrac{1}{4} + it/2\right)\right],$

(2.2) $S(t) = \pi^{-1}\arg\zeta\left(\tfrac{1}{2} + it\right),$

where in both cases the argument is defined by continuous variation of s in $\pi^{-s/2}\Gamma(s/2)$ or $\zeta(s)$, respectively, starting at $s = 2$, going up vertically to $s = 2 + it$, and then horizontally to $s = \frac{1}{2} + it$, and where we assume (in the case of $S(t)$) that there are no zeros ρ with $\text{Im}(\rho) = t$. If $N(t)$ denotes the number of zeros ρ with $0 < \text{Im}(\rho) < t$ (counted according to their multiplicity), then [68, Chapter 9.3]

(2.3) $N(t) = 1 + \pi^{-1}\theta(t) + S(t).$

By Stirling's formula [40] we have

$$(2.4) \qquad \theta(t) = \tfrac{1}{2}t \log(t/(2\pi e)) - \pi/8 + O(t^{-1}) \quad \text{as } t \to \infty,$$

so that

$$(2.5) \qquad N(t) = \frac{t}{2\pi} \log \frac{t}{2\pi e} + \frac{7}{8} + O(t^{-1}) + S(t).$$

We will discuss the function $S(t)$ at greater length below. For the moment, though, we will mention only that the best unconditional bound that is known for $S(t)$ is [68, Theorem 9.4]

$$(2.6) \qquad S(t) = O(\log t) \quad \text{as } t \to \infty.$$

In view of the bound (2.6), we see by (2.5) that $N(t)$ is almost exactly equal to $\pi^{-1}\theta(t)$. In particular, zeros become denser and denser the higher up one goes in the critical strip, and the average vertical spacing between consecutive zeros at height t is $2\pi/(\log(t/(2\pi)))$. Therefore, in order to study quantities that are largely invariant with height, we define the normalized spacing between consecutive zeros $\tfrac{1}{2} + i\gamma_n$ and $\tfrac{1}{2} + i\gamma_{n+1}$ (in cases where both zeros satisfy the RH) to be

$$(2.7) \qquad \delta_n = (\gamma_{n+1} - \gamma_n) \frac{(\log \gamma_n)/(2\pi)}{2\pi}.$$

It then follows from (2.5) and (2.6) that the δ_n have mean value 1 in the sense that for any positive integers N and M,

$$(2.8) \qquad \sum_{n=N+1}^{N+M} \delta_n = M + O(\log NM).$$

In this notation, Montgomery's pair correlation conjecture can be reformulated to say that for any fixed α, β with $0 < \alpha < \beta < \infty$,

$$(2.9) \qquad \begin{aligned} N^{-1} \big| \{(n,k)\colon 1 &\leqslant n \leqslant N,\ k \geqslant 0,\ \delta_n + \delta_{n+1} + \cdots + \delta_{n+k} \in [\alpha, \beta]\} \big| \\ &\sim \int_\alpha^\beta \left(1 - \left(\frac{\sin \pi u}{\pi u}\right)^2\right) du \end{aligned}$$

as $N \to \infty$. If the pair correlation conjecture holds, we might even expect something stronger to hold, namely that

$$(2.10) \qquad \begin{aligned} M^{-1} \big| \{(n,k)\colon N+1 &\leqslant n \leqslant N+M,\ k \geqslant 0,\ \delta_n + \cdots + \delta_{n+k} \in [\alpha, \beta]\} \big| \\ &\sim \int_\alpha^\beta \left(1 - \left(\frac{\sin \pi u}{\pi u}\right)^2\right) du \end{aligned}$$

as $N, M \to \infty$ with M not too small compared to N, say $M \geqslant N^\eta$ for some $\eta > 0$.

The definition of the δ_n makes it easy to compare distribution of spacings between consecutive zeros of the zeta function and the normalized spacings between consecutive eigenvalues in the GUE, since the latter are also normalized to have mean value 1. The *GUE hypothesis* will be used to refer to the conjecture that the distribution of the δ_n approaches asymptotically the distribution known to hold for the GUE eigenvalues. In particular, the GUE hypothesis leads to the prediction that the δ_n should have a particular distribution function; for any α, β with $0 \leqslant \alpha < \beta < \infty$,

$$(2.11) \qquad M^{-1} \big| \{n\colon N+1 \leqslant n \leqslant N+M,\ \delta_n \in [\alpha, \beta]\} \big| \sim \int_\alpha^\beta p(0, u)\, du$$

as M, $N \to \infty$ with M not too small compared to N, where $p(0, u)$ is a certain complicated probability density function discussed in Section 6 and graphed (with the solid line curve) in Figures 3 and 4. Similarly, one obtains the prediction that

$$(2.12) \quad M^{-1}|\{n: N + 1 \leqslant n \leqslant N + M, \delta_n + \delta_{n+1} \in [\alpha, \beta]\}| \sim \int_\alpha^\beta p(1, u) \, du,$$

where $p(1, u)$ is another density function. More generally, one obtains the prediction that the δ_n approach asymptotically the behavior of a stationary (but non-Markovian) process, so that for any k, the empirical joint distribution function of δ_n, $\delta_{n+1}, \ldots, \delta_{n+k}$ for $N + 1 \leqslant n \leqslant N + M$ approaches that of the GUE process.

Most of the studies to be reported below are based on the δ_n for $1 \leqslant n \leqslant 10^5$ and $10^{12} + 1 \leqslant n \leqslant 10^{12} + 10^5$, although a few other sets of 10^5 consecutive δ_n for $n < 10^{12}$ will also be referred to occasionally. The letter N will be used to designate the starting point of a data set, so that, for example, a graph or a table entry labelled with $N = 10^{11}$ will refer to data for δ_n or $\delta_n + \delta_{n+1}$, $10^{11} + 1 \leqslant n \leqslant 10^{11} + 10^5$. The reason for selecting sets of size 10^5 was to obtain a sample so large that random sampling errors would not be very significant. The reason for not computing zeros higher than ρ_n for $n = 10^{12} + 10^5$ (approximately) is that on the machine that was available to the author (a single-processor Cray X-MP with 2 million words of memory) such computations would have been prohibitively slow, since the author's program, described in Section 3, achieves high speed by using several auxiliary storage arrays, so that the size of available memory was a major limit.

Most of the studies of the available δ_n are concerned with the pair correlation conjecture (2.10) and the conjectured distributions (2.11) and (2.12) for δ_n and $\delta_n + \delta_{n+1}$, respectively. The reason for this is less the difficulty of computing the GUE predictions for other quantities than the question of significance of the comparison of the experimental data to the theoretical predictions. By (2.5) and (2.6) we see that the interesting behavior of spacings between zeros is due to $S(t)$. The bound (2.6) is the best that is known unconditionally. Even on the assumption of the RH, it is only known that

$$(2.13) \qquad\qquad S(t) = O\left(\frac{\log t}{\log \log t}\right) \quad \text{as } t \to \infty.$$

The true rate of growth is likely to be much smaller, though. It has been known [68, Chapter 14.12] for a long time that on the RH,

$$S(t) = \Omega_\pm\left((\log t)^{1/2 - \varepsilon}\right) \quad \text{as } t \to \infty$$

for any $\varepsilon > 0$, and Montgomery [58] has shown more recently that

$$(2.14) \qquad\qquad S(t) = \Omega_\pm\left((\log t)^{1/2}(\log \log t)^{-1/2}\right) \quad \text{as } t \to \infty.$$

It is thought likely that

$$(2.15) \qquad\qquad S(t) = O\left((\log t)^{1/2 + \varepsilon}\right) \quad \text{as } t \to \infty$$

for every $\varepsilon > 0$ (cf. [49]) and Montgomery has even conjectured [58] that (2.14) represents the right rate of growth of $S(t)$. For statistical studies of the kind we are undertaking, though, the typical size of $S(t)$ is more important. Selberg [66], [68] has

shown that for all $k \in \mathbf{Z}^+$,

$$(2.16) \qquad \int_0^T S(t)^{2k} \, dt \sim \frac{(2k)!}{k!(2\pi)^{2k}} T(\log \log T)^k \quad \text{as } T \to \infty.$$

This means that a typical value of $S(t)$ is on the order of $(\log \log t)^{1/2}$, an extremely small quantity. Experimentally, it is known that $|S(t)| < 1$ for $7 < t \leqslant 280$, and $|S(t)| < 2$ for $7 < t \leqslant 6.82 \cdot 10^6$. The largest value of $|S(t)|$ that has been observed so far appears to be $S(t) = 2.3136 \cdots$ for some t near γ_n, $n = 1,333,195,692$ [52]. (Large values of $S(t)$ are associated with violations of Gram's and Rosser's rules, see [10], [23], [49], [59], which explains why some statistics about them are available.)

Aside from (2.6) and Selberg's estimate (2.16), several other estimates have been proved. For example, if

$$S_1(t) = \int_0^t S(u) \, du,$$

then $|S_1(t)| = O(\log t)$ unconditionally, and $|S_1(t)| = O((\log t)(\log \log t)^{-2})$ on the RH [68]. The true maximal order of magnitude is probably again around $(\log t)^{1/2}$. Therefore, there is a lot of cancellation in the local behavior of $S(t)$. Many results about the distribution of values of $S(t + h) - S(t)$ are also known [25], [26], [27], [32], [33], [44], [69]. For example, Theorem 13.2 of [69] says that if $\frac{1}{2} < \eta \leqslant 1$, then for $T^\eta \leqslant H \leqslant T, 0 < h < 1$,

$$(2.17) \qquad \begin{aligned} &\int_T^{T+H} \{ S(t + h) - S(t) \}^{2k} \, dt \\ &= HA_k \{ \log(2 + h \log T) \}^k + O\Big(Hc^k k^k \big\{ k^k + (\log(2 + h \log T))^{k-1/2} \big\} \Big), \end{aligned}$$

where $A_k = (2k)!/(2^k \pi^{2k} k!)$, and $c > 0$ is a constant. There are also theorems about the distribution of $S_1(t + h) - S_1(t)$; cf. [69].

Since $S(t)$ grows very slowly, we can expect low zeros of the zeta function to be very restricted in their behavior, and their asymptotic behavior to be approached quite slowly. Pseudo-random behavior of the zeros (like that predicted by the GUE hypothesis) can only be expected over ranges of about $(\log T)^{-1/2}$ at height T; i.e., over a range of about $(\log n)^{1/2}$ normalized spacings from zero number n. Furthermore, we can expect gaps between next-to-nearest neighbors (i.e., $\delta_n + \delta_{n+1}$) to be much more constrained than those between nearest neighbors (i.e., δ_n), and to approach their asymptotic behavior even more slowly. This is indeed what is observed, and it is the main reason for restricting most of the present investigation to nearest or next-to-nearest spacings. We will see, in fact, that when we consider very distant spacings, a totally different phenomenon occurs, whose explanation lies not in the GUE, but rather in prime numbers.

We now proceed to a discussion of the evidence. Figures 1 and 2 show the data for the pair correlation conjecture for the two sets of zeros $\frac{1}{2} + i\gamma_n$, $N + 1 \leqslant n \leqslant N + 10^5$, and for $N = 0$ and $N = 10^{12}$, respectively. For each interval $I = [\alpha, \beta)$, with $[\alpha, \beta) = [0, 0.05), [0.05, 0.1), \ldots, [2.95, 3)$, a star is plotted at the point $x = (\alpha + \beta)/2$, $y = a_{\alpha,\beta}$, where

$$a_{\alpha,\beta} = \frac{20}{10^5} \Big| \big\{ (n, k) : N + 1 \leqslant n \leqslant N + 10^5, \ k \geqslant 0, \ \delta_n + \cdots + \delta_{n+k} \in [\alpha, \beta) \big\} \Big|.$$
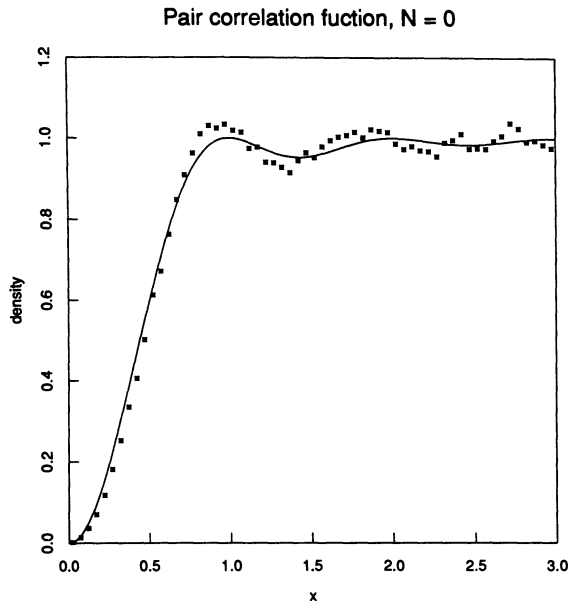
FIGURE 1

Pair correlation of zeros of the zeta function. Solid line: GUE prediction. Scatter plot: empirical data based on zeros γ_n, $1 \leqslant n \leqslant 10^5$.
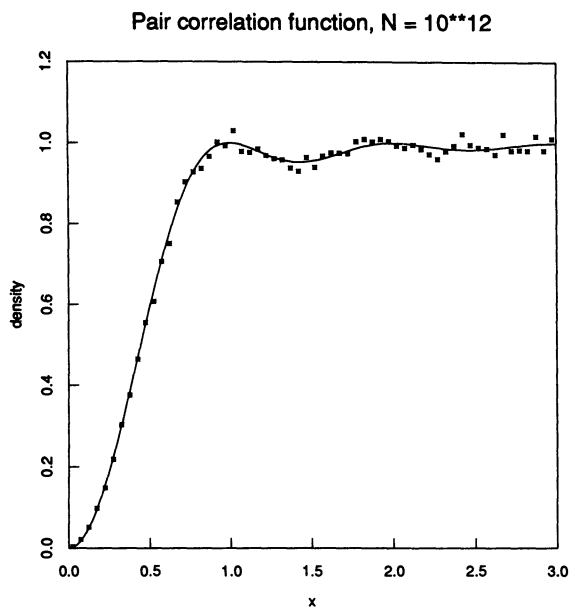

FIGURE 2

Pair correlation of zeros of the zeta function. Solid line: GUE prediction. Scatter plot: empirical data based on zeros γ_n, $10^{12} + 1 \leqslant n \leqslant 10^{12} + 10^5$.

The solid lines in both figures are the GUE prediction $y = 1 - ((\sin \pi x)/(\pi x))^2$. Note that for a typical value of $\alpha \geqslant 1$, $a_{\alpha, \beta}$ is derived from about 5000 points $\delta_n + \cdots + \delta_{n+k}$, so the random sampling error is expected to be on the order of 0.015. All of the plots in this paper, as well as most of the statistical computations, were done using the S system [1].

Figures 3 and 4 show the data on the distribution of the normalized spacings δ_n. The stars again correspond to a histogram of the δ_n; for each interval $[\alpha, \beta)$, $\alpha = k/20$, $\beta = \alpha + 1/20$, a star is plotted at $x = (\alpha + \beta)/2$, $y = b_{\alpha, \beta}$, where

$$b_{\alpha, \beta} = \frac{20}{10^5} \left| \left\{ n \colon N + 1 \leqslant n \leqslant N + 10^5, \, \delta_n \in [\alpha, \beta) \right\} \right|.$$

The solid lines are the GUE predictions, $y = p(0, x)$. (As is explained in Section 6, no rigorous error analysis is available for the computed values of $p(0, x)$, but they are thought to be quite accurate.) Similarly, Figures 5 and 6 show the data on the distribution of $\delta_n + \delta_{n+1}$.

The graphs at first glance show a satisfying amount of agreement between the experimental data and the GUE predictions. Graphs have roughly the same shape, and agreement between data and prediction improves dramatically as one goes from the first 10^5 zeros to the 10^5 zeros with $10^{12} + 1 \leqslant n \leqslant 10^{12} + 10^5$. Moreover, the disagreement between data and prediction is concentrated precisely where our discussion of the effect of the small size of $S(t)$ would lead one to expect it; there are fewer very large or very small values of δ_n and $\delta_n + \delta_{n+1}$ in the data than predicted, and more values near the mean.

Nearest neighbor spacings, N = 0
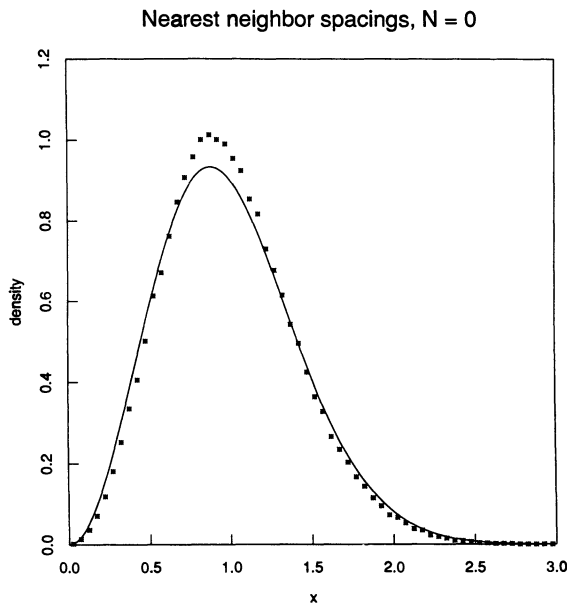
FIGURE 3

Probability density of the normalized spacings δ_n. *Solid line: GUE prediction. Scatter plot: empirical data based on zeros* γ_n, $1 \leqslant n \leqslant 10^5$.

A. M. ODLYZKO

Nearest neighbor spacings, N = 10**12
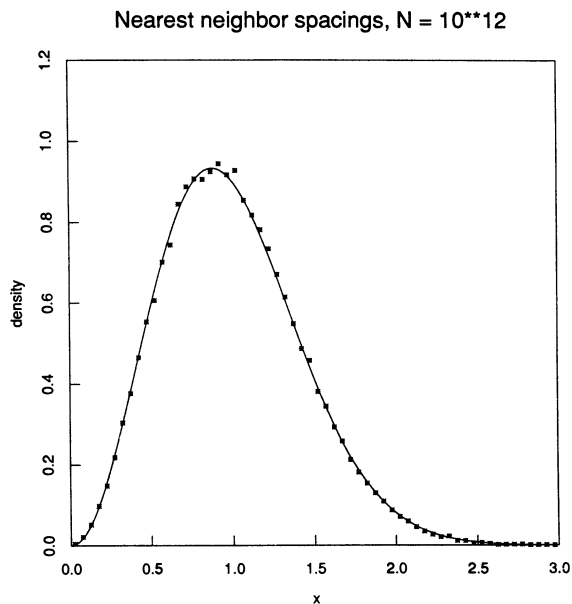
FIGURE 4

Probability density of the normalized spacings δ_n. *Solid line:*
GUE prediction. Scatter plot: empirical data based on zeros
γ_n, $10^{12} + 1 \leqslant n \leqslant 10^{12} + 10^5$.

Next-nearest neighbor spacings, N = 0

FIGURE 5

Probability density of the normalized spacings $\delta_n + \delta_{n+1}$. *Solid*
line: GUE predition. Scatter plot: empirical data based on
zeros γ_n, $1 \leqslant n \leqslant 10^5$.
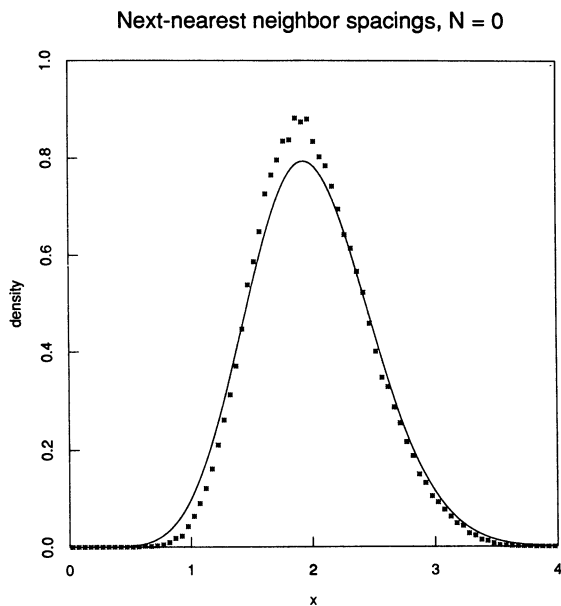
Next-nearest neighbor spacings, N = 10**12
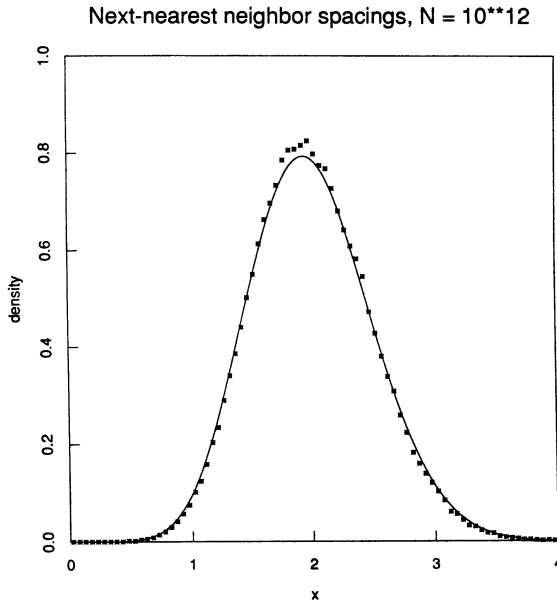
FIGURE 6

Probability density of the normalized spacings $\delta_n + \delta_{n+1}$. Solid line: GUE prediction. Scatter plot: empirical data based on zeros γ_n, $10^{12} + 1 \leqslant n \leqslant 10^{12} + 10^5$.

TABLE 1

Moments of $\delta_n - 1$ and of $\delta_n + \delta_{n+1} - 2$, for blocks of the form $N + 1 \leqslant n \leqslant N + 10^5$

k	$\delta_n - 1$			$\delta_n + \delta_{n+1} - 2$		
	$N = 0$	$N = 10^{12}$	GUE	$N = 0$	$N = 10^{12}$	GUE
2	0.161	0.176	0.180	0.207	0.236	0.249
3	0.031	0.035	0.038	0.028	0.027	0.030
4	0.081	0.096	0.101	0.123	0.168	0.185
5	0.046	0.057	0.066	0.047	0.063	0.073
6	0.075	0.098	0.111	0.119	0.206	0.237
7	0.072	0.103	0.124	0.078	0.150	0.185
8	0.103	0.160	0.197	0.155	0.367	0.451
9	0.126	0.223	0.290	0.142	0.409	0.544
10	0.181	0.361	0.488	0.252	0.871	1.178

Further comparison of the distribution of zeros to the GUE prediction is provided by Table 1, which shows values of

$$(2.18) \qquad 10^{-5} \sum_{n=N+1}^{N+10^5} \left(\delta_n - 1 \right)^k$$

and

(2.19) $$10^{-5} \sum_{n=N+1}^{N+10^5} (\delta_n + \delta_{n+1} - 2)^k$$

for the two sets $N = 0$ and $N = 10^{12}$, as well as the GUE values. (Numbers in Table 1, as well as others in this paper that do not have " $\cdots$ " at the end are usually rounded to the number of digits shown, whereas those with " $\cdots$ " are truncated.) Table 2 shows moments of $\log \delta_n$, δ_n^{-1}, and δ_n^{-2}.

TABLE 2

Moments of $\log \delta_n$, δ_n^{-1}, *and* δ_n^{-2}, *for blocks of the form*
$N + 1 \leqslant n \leqslant N + 10^5$.

moments of	$N = 0$	$N = 10^{12}$	GUE
$\log \delta_n$	–0.0912	–0.1021	–0.1035
δ_n^{-1}	1.2363	1.2744	1.2758
δ_n^{-2}	2.2235	2.6149	2.5633

The histograms of Figures 1–6 and the moments of Tables 1–2 provide a very rough measure of the degree to which zeros of the zeta function match the GUE predictions. One can obtain a better quantitative measure of the degree of fit between the observed and the predicted distributions using the histogram data (e.g., one can use the χ^2-test [47] or the asymptotic results of [24]). In our case, since we know the predicted distribution quite well (at least numerically), we will use the quantile-quantile (q-q) plot to detect deviations between the empirical and theoretical distributions. Given a sample $x_1, \ldots, x_n$, and a continuous cumulative distribution function $F(z)$ for some distribution, the theoretical q-q plot is obtained by plotting $x_{(j)}$ against q_j, where $x_{(1)} \leqslant x_{(2)} \leqslant \cdots \leqslant x_{(n)}$ are the x_i sorted in ascending order, and the q_j are the theoretical quantiles defined by $F(q_j) = (j - 1/2)/n$ [13]. The q-q plot is a very sensitive tool for detecting differences among distributions, especially near the tails. If the x_i are drawn from a distribution corresponding to $F(z)$, and the sample size n is large, the q-q plot will be very close to the straight line $y = x$. Figures 7–10 show segments of the q-q plots of the δ_n and $\delta_n + \delta_{n+1}$ for $N = 0$ and $N = 10^{12}$.

To obtain a quantitative measure of the agreement between the observed distributions of the δ_n and the GUE predictions, we use the Kolmogorov test [47, Section 30.49]. Given a sample $x_1, \ldots, x_n$ drawn from a distribution with the continuous cumulative distribution function $F(z)$, we let $F_e(z)$ be the sample distribution function:

$$F_e(z) = n^{-1} | \{ k : 1 \leqslant k \leqslant n, x_k \leqslant z \} |.$$

The Kolmogorov statistic is then

(2.20) $$D = \sup_z |F_e(z) - F(z)|.$$

FIGURE 7

Initial segment of the quantile-quantile plot of the normalized spacings δ_n against the GUE prediction. Data based on zeros γ_n, $1 \leqslant n \leqslant 10^5$. Straight line $y = x$ drawn to facilitate comparisons.
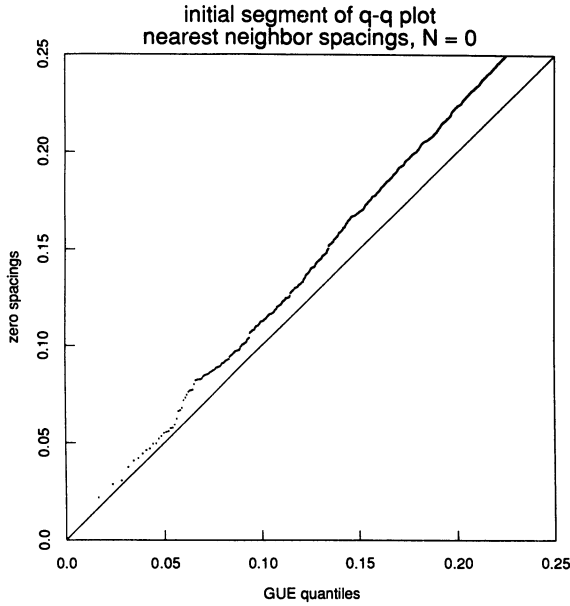

FIGURE 8

Initial segment of the quantile-quantile plot of the normalized spacings δ_n against the GUE prediction. Data based on zeros γ_n, $10^{12} + 1 \leqslant n \leqslant 10^{12} + 10^5$. Straight line $y = x$ drawn to facilitate comparisons.

final segment of q-q plot
next-nearest neighbor spacings, N = 0

FIGURE 9

Final segment of the quantile-quantile plot of the normalized spacings $\delta_n + \delta_{n+1}$ against the GUE prediction. Data based on zeros γ_n, $1 \leqslant n \leqslant 10^5$. Straight line $y = x$ drawn to facilitate comparisons.

final segment of q-q plot
next-nearest neighbor spacings, N = 10**12

FIGURE 10

Final segment of the quantile-quantile plot of the normalized spacings $\delta_n + \delta_{n+1}$ against the GUE prediction. Data based on zeros γ_n, $10^{12} + 1 \leqslant n \leqslant 10^{12} + 10^5$. Straight line $y = x$ drawn to facilitate comparisons.

The asymptotic distribution of D is known [47, Eq. 30.132]; if the x_i are drawn from the distribution defined by $F(z)$, then

$$(2.21) \qquad \lim_{n \to \infty} \text{Prob}(D > un^{-1/2}) = g(u),$$

where

$$(2.22) \qquad g(u) = 2\sum_{r=1}^{\infty} (-1)^{r-1}\exp(-2r^2u^2).$$

Table 3 gives the Kolmogorov statistic D for δ_n and $\delta_n + \delta_{n+1}$ for several blocks of 10^5 consecutive zeros. Also given is an estimate of the probability that this statistic would arise if the δ_n ($\delta_n + \delta_{n+1}$, respectively) were drawn from the GUE distribution. This estimate is obtained by evaluating $g(D\,10^{5/2})$.

TABLE 3

Kolmogorov statistic for δ_n and $\delta_n + \delta_{n+1}$, for blocks of zeros of the form $N + 1 \leqslant n \leqslant N + 10^5$.

	δ_n		$\delta_n + \delta_{n+1}$	
N	D	prob.	D	prob.
0	0.0207	10^{-37}	0.0278	10^{-67}
10^9	0.0081	$4 \cdot 10^{-6}$	0.0138	10^{-16}
10^{11}	0.0052	$9 \cdot 10^{-3}$	0.0099	$6 \cdot 10^{-9}$
$2 \cdot 10^{11}$	0.0058	$2 \cdot 10^{-3}$	0.0084	$1.5 \cdot 10^{-6}$
10^{12}	0.0045	$3.5 \cdot 10^{-2}$	0.0091	$1.3 \cdot 10^{-6}$

The data presented so far are fairly consistent with the GUE predictions. The differences between computed values and the predicted ones diminish as one studies higher zeros, and they are more pronounced for $\delta_n + \delta_{n+1}$ than for δ_n. Moreover, the computed spacings δ_n and $\delta_n + \delta_{n+1}$ are more concentrated near their means than the GUE predictions, which is again to be expected on the basis of the small size of $S(t)$. However, there are some indications, based on the tails of the distribution of the δ_n and $\delta_n + \delta_{n+1}$, that asymptotically the zeros of the zeta function might not obey the GUE predictions. Already in Table 2 we see that the mean value of δ_n^{-2} for $N = 10^{12}$ exceeds the GUE prediction, which indicates an excess of small δ_n. The q-q plot of Figure 8 (for $N = 10^{12}$) is much closer to a straight line than that of Figure 7 (for $N = 0$), just as that of Figure 10 is closer to a straight line than that of Figure 9. However, the graphs in Figures 9 and 10 are both below the straight line $y = x$ (which corresponds to perfect agreement between theory and prediction), which indicates that in both cases there are fewer large spacings than the GUE hypothesis predicts, which was to be expected. The graph of Figure 7 is above the straight line $y = x$, which indicates fewer small spacings than predicted, which again was to be expected, while on the other hand the graph of Figure 8 is below the line $y = x$, which indicates an excess of small spacings, which is quite unexpected. The small size of $S(t)$ leads one to expect a substantial deficiency in the number of small δ_n

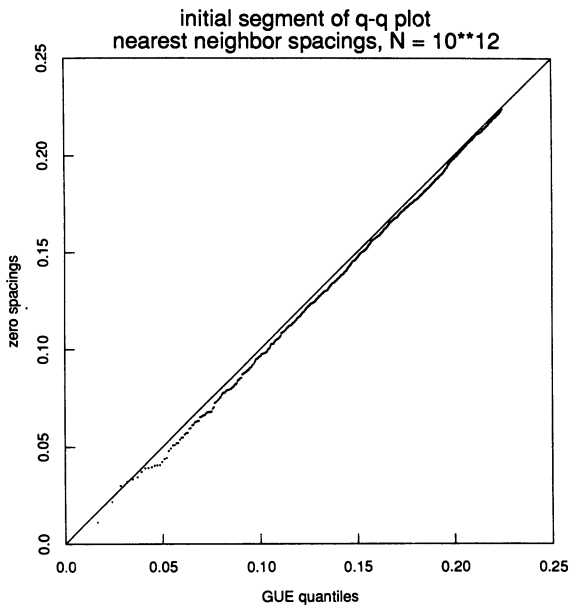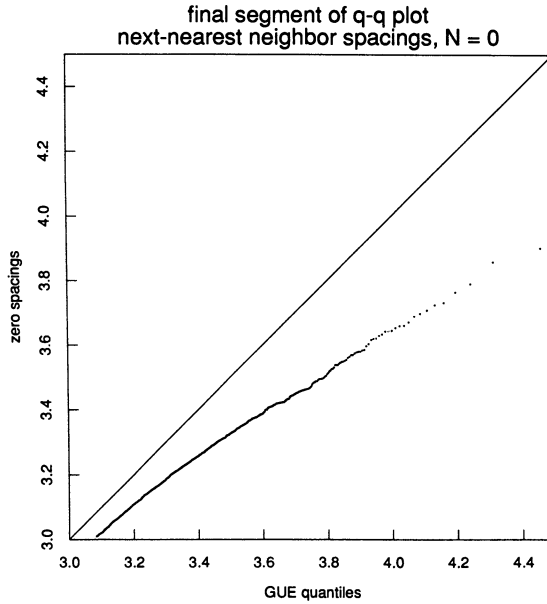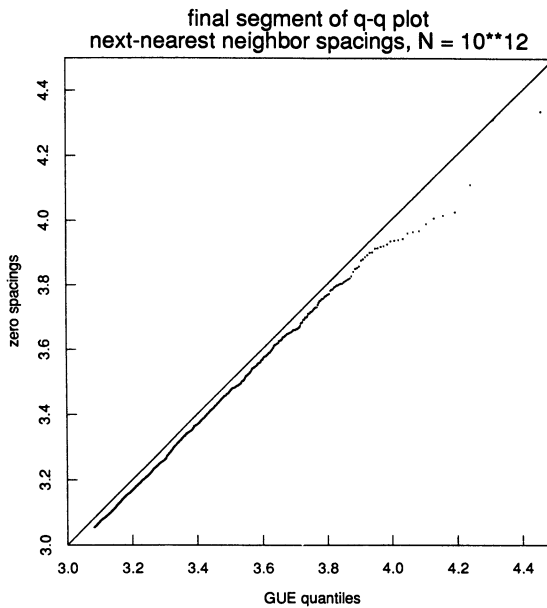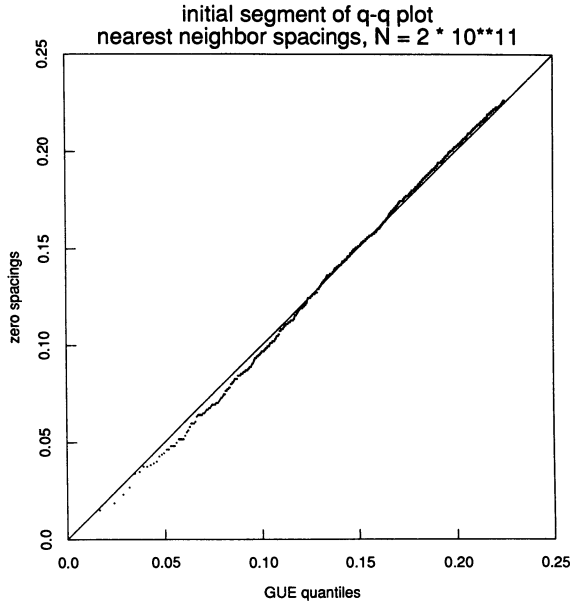initial segment of q-q plot
nearest neighbor spacings, N = 2 * 10**11

FIGURE 11

Initial segment of the quantile-quantile plot of the normalized spacings δ_n against the GUE prediction. Data based on zeros γ_n, $2 \cdot 10^{11} + 1 \leqslant n \leqslant 2 \cdot 10^{11} + 10^5$. Straight line $y = x$ drawn to facilitate comparisons.

and $\delta_n + \delta_{n+1}$, since the graph of $S(t)$ is an uneven sawtooth curve, with $S(t)$ jumping up by 1 at $t = \gamma_n$, then decreasing essentially linearly up to $t = \gamma_{n+1}$, jumping up again by 1 at $t = \gamma_{n+1}$, and so on, so that small gaps between zeros also correspond to large values of $S(t)$. Figure 11 shows an initial part of the q-q plot for the δ_n for $N = 2 \cdot 10^{11}$ which indicates behavior very similar to that of Figure 8. The q-q plot for $N = 10^{11}$ is very similar to those for $N = 2 \cdot 10^{11}$ and $N = 10^{12}$, while that of $N = 10^9$ lies on the other side of the $y = x$ line. (We should note again that the data for $N = 10^9$, 10^{11}, and $2 \cdot 10^{11}$ are not very accurate, and so have to be treated with caution. In fact, Figure 11 provides a way to gauge the inaccuracy introduced into those data sets by the bug described in Section 3. The q-q plot of Figure 11 has in places the appearance of a staircase, with flat horizontal segments. These flat segments arise from identical computed values of the δ_n. The correct values of the δ_n are expected to be all distinct and much more evenly distributed, as is the case for $N = 0$ and $N = 10^{12}$.) Table 4 shows the number of δ_n and $\delta_n + \delta_{n+1}$ that are very small or very large, as well as the number that would be expected under the GUE for a sample of size 10^5. Large δ_n and $\delta_n + \delta_{n+1}$ are generally less numerous than the GUE predicts, which is to be expected. The data for the high blocks of zeros show that the number of $\delta_n \leqslant 0.05$ and of $\delta_n \leqslant 0.1$ is in excess of that predicted. (We would have obtained the same results if we considered the pair correlation conjecture (2.10) with $\alpha = 0.01$, $\beta = 0.05$, say.) The excess is not very large, given that the expected number is fairly small, but it may be significant that while the first few sets (for $N = 0$, $5 \cdot 10^7$, and 10^9) all show deficiencies, all the high sets of zeros ($N = 10^{11}$, $2 \cdot 10^{11}$, and 10^{12}) show an excess of $\delta_n \leqslant 0.05$ and of $\delta_n \leqslant 0.1$.

TABLE 4

Numbers of very small and very large δ_n and $\delta_n + \delta_{n+1}$ in blocks of the form $N + 1 \leqslant n \leqslant N + 10^5$, and the GUE predictions.

	number of $\delta_n \leq 0.05$	number of $\delta_n \leq 0.1$	number of $\delta_n \geq 2.8$	number of $\delta_n + \delta_{n+1} \leq 0.6$	number of $\delta_n + \delta_{n+1} \geq 4$
$N = 0$	11	79	1	2	0
$N = 5 \cdot 10^7$	9	103	9	24	3
$N = 10^9$	11	97	14	33	5
$N = 10^{11}$	16	107	11	41	8
$N = 2 \cdot 10^{11}$	20	123	20	25	9
$N = 10^{12}$	18	124	14	34	6
GUE	13.68	108.80	19.68	38.63	13.57

TABLE 5

Extremal values of δ_n and $\delta_n + \delta_{n+1}$ among zeros number n, $N + 1 \leqslant n \leqslant N + M$, and the probability that the minimum value of δ_n in the GUE would not exceed the value in the third column.

N	M	min δ_n	max δ_n	min $\delta_n + \delta_{n+1}$	max $\delta_n + \delta_{n+1}$	prob. min δ_n
0	10^5	0.02186	2.8052	0.57647	3.9022	0.682
$5 \cdot 10^7$	10^5	0.00734	3.1551	0.41841	4.3058	0.042
10^9	10^5	0.01679	3.0680	0.37404	4.1650	0.405
10^{11}	10^5	0.00738	3.2811	0.30784	4.4753	0.043
$2 \cdot 10^{11}$	10^5	0.01515	3.1109	0.39744	4.3121	0.317
10^{12}	10^5	0.01103	3.1893	0.44260	4.3371	0.137
0	10^6	0.00970	3.2143	0.26344	4.1517	0.632
0	$7.5 \cdot 10^7$	≤ 0.00124				≤ 0.145
0	$1.5 \cdot 10^9$	≤ 0.000310				≤ 0.048

Another way to investigate this question is to look at the minimal values of δ_n, shown in Table 5. These values include some very small δ_n that were found by Brent [10] and van de Lune et al. [52]. These investigators were not computing the γ_n, but occasionally, when they had difficulty separating a close pair of zeros, they noted the approximate locations of them. In particular, Brent [10] found that $\delta_n = 0.001235 \cdots$ for $n = 41,820,581$, and van de Lune et al. [52] found that $\delta_n = 0.000310 \cdots$ for $n = 1,048,449,114$. Thus, these values provide upper bounds for the minimal δ_n in the intervals investigated by those authors, and it is likely that they are the minimal values.

Since it is known [54], [55] that $p(0, t)$ has the Taylor series expansion near $t = 0$ given by

$$(2.23) \qquad p(0, t) = \frac{\pi^2}{3} t^2 - \frac{2\pi^4}{45} t^4 + \frac{\pi^6}{315} t^6 + \cdots,$$

(for comparison, we note that $p(1, t) = \pi^6 t^7 / 4050 + \dots$ [55]), the GUE prediction is that among M consecutive δ_n's, the probability that the minimal δ_n is $\leqslant \alpha M^{-1/3}$ is approximately

$$(2.24) \qquad\qquad 1 - \left(1 - \frac{\pi^2}{9}\frac{\alpha^3}{M}\right)^M \sim 1 - \exp(-\pi^2 \alpha^3 / 9).$$

The function on the right side of (2.24) was used to compute the last column in Table 5. The probabilities of obtaining the minimal δ_n that are observed from the GUE distribution are rather low, in some cases very low. (We would have obtained comparable inconsistencies between data and predictions had we worked with the pair correlation function instead of the distribution of the δ_n.)

It is hard to draw any firm conclusions from the data that is available. It would be very desirable to obtain data from much higher sets of zeros to see whether they also have an excess of smaller spacings compared to the GUE hypothesis. If they do, this would raise doubt not only about the GUE hypothesis but even about the Montgomery pair correlation conjecture. This is an important possibility because the GUE hypothesis is very speculative, as it relies on the hope that the eigenvalues of the hypothetical operator associated with the zeta function would behave like those of a random operator. Thus it is easy to conceive of the GUE hypothesis being false even when there is an appropriate operator. On the other hand, the pair correlation conjecture depends only on the assumption of randomness in the behavior of primes (basically on error terms in the number of primes in arithmetic progressions cancelling each other out). Therefore, falsity of the pair correlation conjecture would imply some unusual behavior among the primes.

It would also be desirable to obtain numerical data for zeros of Dirichlet L-functions, using the generalization of the Riemann-Siegel formula [20], [21] or the new method of [61]. Some data are available for zeros of Epstein zeta functions [8], and they do not obey the GUE predictions, as they have many small spacings between consecutive zeros. This is not surprising, since Epstein zeta functions do not satisfy the RH in general.

We briefly discuss large δ_n. By [15],

$$(2.25) \qquad\qquad \log p(0, t) \sim -\frac{\pi^2}{8}t^2 \quad \text{as } t \to \infty$$

(a result that had been derived earlier, but less rigorously, by F. J. Dyson [22]). Hence the GUE prediction is that

$$\max_{N+1 \leqslant n \leqslant N+M} \delta_n \sim \pi^{-1}(8 \log M)^{1/2}$$

as N, $M \to \infty$ with M not too small compared with N, and this implies

$$(2.26) \qquad\qquad \max_{\gamma_n < T}(\gamma_{n+1} - \gamma_n) \sim 8(2 \log T)^{-1/2}.$$

Montgomery's conjecture [58, Eq. (11)] that

$$|S(t)| = O\left((\log t)^{1/2}(\log \log t)^{-1/2}\right),$$

as well as another argument of Joyner [45], suggest that

$$(2.27) \qquad \max_{\gamma_n < T} (\gamma_{n+1} - \gamma_n) = O\big((\log T)^{-1/2}(\log \log T)^{-1/2}\big),$$

which contradicts (2.26). Because of the slow growth of $(\log T)^{1/2}$ and of $(\log \log T)^{1/2}$, our data do not allow us to conclude anything about the truth of these conjectures. (Some additional information about known large values of δ_n is presented in Section 4.)

There are some measures of higher-order correlation among zeros that were computed and compared to the GUE prediction [7]. The GUE predictions are not known very accurately, but the agreement seemed quite good for short-range correlations. Substantial deviations were found in the measure $\Sigma^2(r)$ of [7] for large r, which equals the variance of the number of (normalized) zeros in an interval of length r. In the GUE case, $\Sigma^2(r)$ is monotone increasing, whereas for the zeros of the zeta function for $N = 10^{12}$, this measure was increasing up to about $r = 13$, and then started to decrease slightly. Given the bounds on $S(t)$, this is not very surprising.

One feature of the behavior of the zero spacings which initially seemed quite surprising involves very long-range correlations among the δ_n. As is usual, we define the autocovariances of the δ_n by

$$(2.28) \qquad c_k = c_k(N, M) = \frac{1}{M} \sum_{n=N+1}^{N+M} (\delta_n - 1)(\delta_{n+k} - 1).$$

It is not known exactly what the c_k are in the GUE, but it has been conjectured by F. J. Dyson (unpublished) that for $k > 0$,

$$(2.29) \qquad c_k \approx \frac{-1}{2\pi^2 k^2},$$

where $\approx$ in (2.29) indicates approximate rather than asymptotic value. This conjecture is intuitively appealing, since a large spacing is expected to lead to smaller spacings nearby to preserve the average distance, but this effect should apply less and less the further apart the spacings one considers. What one observes in practice, though, is quite different. Table 6 shows the values of the $c_k(N, M)$ for various values of k, $M = 10^5$, and for the two sets corresponding to $N = 0$ and $N = 10^{12}$. Figure 12 shows a graph of $c_k(10^{12}, 10^5)$ for $950 \leqslant k \leqslant 1000$. Since the variance of the δ_n is around $1/6$, random independent δ_n would give values of c_k around $(6\sqrt{10^5})^{-1} \approx 0.00053$. The values in Table 5 are typically much larger than that, and therefore statistically significant. These values say, for example, that a large δ_n tends to be associated with a small δ_{n+963} (for $N = 10^{12}$). Even more striking than the sizes of the c_k are the patterns of signs that are visible in Table 6 and Figure 12. (These effects are much more striking for $N = 10^{12}$ than for $N = 0$ due to the "averaging out" of the spectrum of the δ_n for $N = 0$ that is introduced by the normalization (2.7), but we will not discuss this here.)

The long-range correlations between the δ_n, which will be discussed below, should not obscure the fact that our data do support the Dyson conjecture (2.29), at least in the weak form that says that for every fixed k, as $N, M \to \infty$ with M not too small

TABLE 6

Autocovariances of the δ_n, $N + 1 \leqslant n \leqslant N + 10^5$.

k	$N = 0$	$N = 10^{12}$
0	.1607429	.1762933
1	-.0574023	-.0582498
2	-.0126083	-.0143371
3	-.0065874	-.0047603
4	-.0045317	-.0035270
5	-.0031454	-.0011576
6	-.0011362	-.0015305
7	-.0007084	-.0008964
8	-.0013904	-.0009071
9	.0013483	-.0011673
10	.0034456	-.0001692
11	.0018714	-.0009839
12	-.0002503	-.0002831
13	-.0005412	-.0003967
14	.0025227	-.0011130
15	.0046388	-.0002101
16	.0025451	-.0005048
17	.0010829	.0003850
55	-.0039299	-.0102228
56	-.0013716	-.0069591
57	.0032256	-.0011287
58	.0016816	.0012409
59	-.0014840	.0020236
60	-.0041143	.0017857
61	-.0017425	.0011242
62	.0019009	.0015995
63	.0000127	.0012965
960	.0002534	.0047463
961	-.0002862	.0013039
962	.0003855	-.0039218
963	-.0002286	-.0061988
964	.0001077	-.0015785
965	-.0000796	-.0011387
966	-.0002891	-.0050911
967	.0000783	-.0053232
968	-.0001676	.0008364
969	-.0000848	.0044897

compared to N, c_k is approximated by the quantity on the right side of (2.29). Some support for this conjecture can be derived from Table 6. The fact that (2.29) does not hold for k large is to be expected, given the known results about $S(t)$. The only thing that is surprising is that the c_k are large for large k and there are patterns to their behavior.

An explanation for the long-range correlations between the δ_n can be obtained by looking at the spectrum of the δ_n. Figure 13 shows a portion of the spectrum, i.e., a graph of

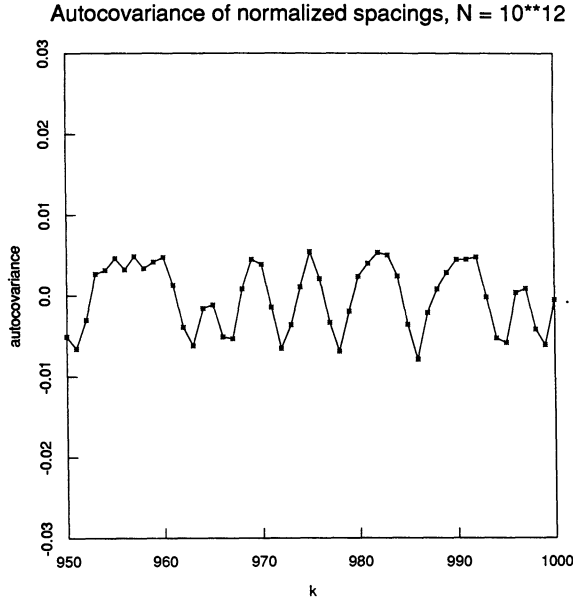$$f(x) = c \left| \sum_{n=N+1}^{N+M} (\delta_n - 1) e^{\pi i n x} \right|^2$$

Autocovariance of normalized spacings, N = 10**12

FIGURE 12

Autocovariance of the δ_n. The entry for k is the mean value of $(\delta_n - 1)(\delta_{n+k} - 1)$, averaged over $10^{12} + 1 \leqslant n \leqslant 10^{12} + 10^5$.

spectrum of normalized differences
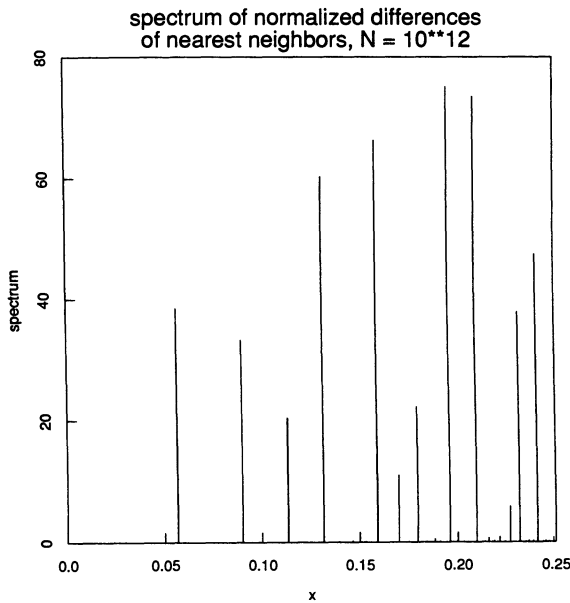of nearest neighbors, N = 10**12

FIGURE 13

Spectrum of the δ_n. Value plotted for x is $c|\sum_n(\delta_n - 1) \cdot \exp(\pi inx)|^2$, where c is a constant, and n runs over $10^{12} + 1 \leqslant n \leqslant 10^{12} + 98303$. (Heights of peaks are not represented accurately due to limited sampling.)
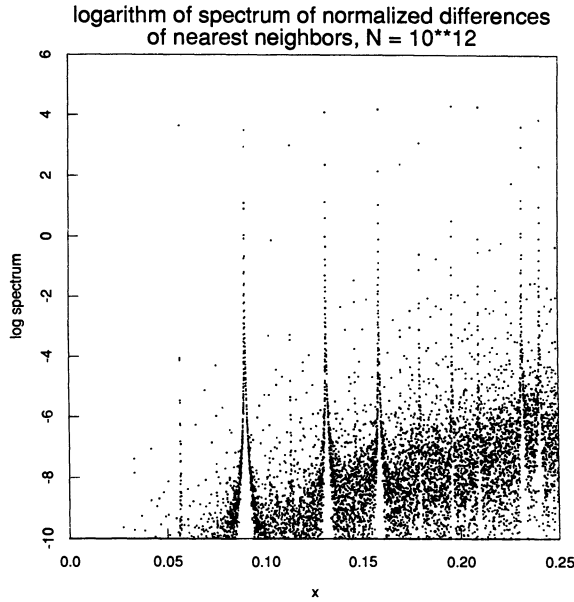
logarithm of spectrum of normalized differences
of nearest neighbors, N = 10**12

FIGURE 14

Logarithm of the spectrum of the δ_n. Value plotted for x is
$2\log|\Sigma_n(\delta_n - 1)\exp(\pi inx)| + c$ *for a constant c, where n runs*
over $10^{12} + 1 \leqslant n \leqslant 10^{12} + 98303$, and where values < -10
are replaced by -10.

for a certain constant c, where $N = 10^{12}$, $M = 98303$. No smoothing or tapering
was done to the data. The spectral lines are very sharp. The line graph of Figure 13
is based on approximately 12000 evenly-spaced values of x between 0 and $1/4$, so
that the heights of the peaks may not be very accurately portrayed, but their
locations are accurate. (The function $f(x)$ is periodic with period 1. For $0.25 \leqslant x \leqslant 1$,
it has more sharp peaks, but they get closer and closer together and eventually begin
to coalesce into a chaotic regime.) To show more clearly the behavior of the
spectrum of the δ_n, Figure 14 shows a graph of $g(x) = \max(\log f(x), -10)$. (This
definition is meant to deemphasize regions where $f(x)$ is small.)

The peaks of $f(x)$ occur near $x = 2(\log N)^{-1}\log q$, where q is a prime power, and
the spikes in Figure 13 are due to 2, 3, 4, 5, 7, 8, 9, 11, 13, 16, 17, and 19, in that
order, with proper prime powers having smaller peaks than primes. The explanation
for this behavior of the spectrum of the δ_n lies in the formulas that connect zeros of
the zeta function and prime numbers. Some of them, such as the exact formula for
$\pi(x)$ in terms of the zeros, or the more general "explicit formulas" of Guinand [38]
and Weil [72] are well known. A related formula was proved by Landau [48]. (See
[37] for a somewhat stronger version.) It says, under the assumption of the RH
(which we assume for convenience, although Landau proved an unconditional
result), that for any $y > 0$ as $N \to \infty$,

$$\sum_{n=1}^{N} e^{i\gamma_n y} = \begin{cases} -\dfrac{\gamma_N}{2\pi}\exp(-y/2)\log p + O(\exp(-y/2)\log N) & \text{if } y = \log p^m, \\ O(\exp(-y/2)\log N) & \text{if } y \neq \log p^m, \end{cases}$$
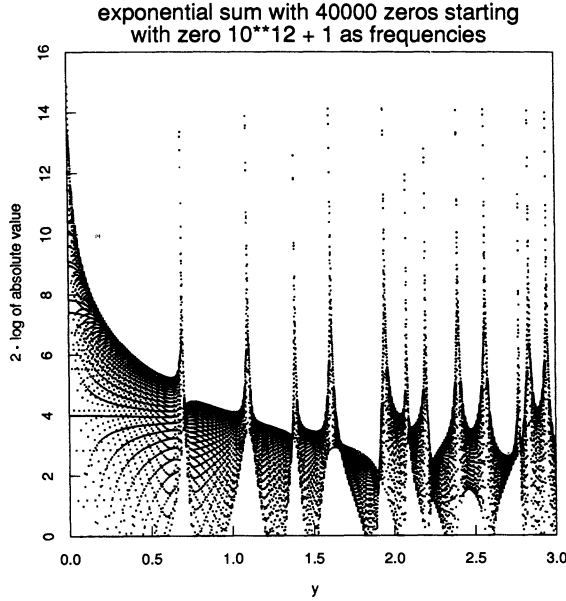
exponential sum with 40000 zeros starting
with zero 10**12 + 1 as frequencies

FIGURE 15

Graph of $2\log|\Sigma \exp(i\gamma_n y)|$ *for* $0.004 \leqslant y \leqslant 3$, *where n runs over* $10^{12} + 1 \leqslant n \leqslant 10^{12} + 40000$, *and values* < 0 *are replaced by* 0.

where p denotes a prime and m a positive integer. Therefore, we also expect that if

$$h(y) = \sum_{n=N+1}^{N+M} e^{i\gamma_n y},$$

then $h(y)$ will be large precisely for y in the vicinity of $\log p^m$ and small elsewhere. Figure 15 shows a graph of $2 \log|h(y)|$ for $0.004 \leqslant y \leqslant 3$, $N = 10^{12}$, $M = 4 \cdot 10^4$, which exhibits precisely this behavior. (This function $2 \log|h(y)|$ is not graphed for $0 \leqslant y < 0.004$, since it is very large there.) Let us write

$$\gamma_{N+k} = \gamma_N + k\alpha + \beta_k,$$

where α is the average spacing,

$$\alpha = 2\pi \left(\log \frac{\gamma_N}{2\pi} \right)^{-1}.$$

The β_k are usually small, on the order of $1/\alpha$. Then, for y small, we can expect that

$$h(y) = e^{i\gamma_N y} \sum_{k=1}^{M} e^{ik\alpha y + i\beta_k y}$$

(2.30)

$$\approx e^{i\gamma_N y} \sum_{k=1}^{M} e^{ik\alpha y} + iye^{i\gamma_N y} \sum_{k=1}^{M} \beta_k e^{ik\alpha y}.$$

The first sum on the right side of (2.30) is small for y away from integer multiples of $2\pi/\alpha$, and so it is the second sum that contributes the oscillations at $y = \log p^m$. (For y away from $\log p^m$, and especially so for y small, the first sum on the right

side of (2.30) dominates, and since it equals the rapidly oscillating function $|\sin(M\alpha y/2)|\ |\sin(\alpha y/2)|^{-1}$ in absolute magnitude, it gives rise to the very regular patterns of points visible in Figure 15 because of the comparatively infrequent but regular spacing of sample points.) On the other hand,

$$\sum_{n=N+1}^{N+M} (\delta_n - 1) e^{\pi i n x} \approx \alpha^{-1} \sum_{n=N+1}^{N+M} (\gamma_{n+1} - \gamma_n - \alpha) e^{\pi i n x}$$

$$= \alpha^{-1} e^{\pi i N x} \sum_{k=1}^{M} (\beta_{k+1} - \beta_k) e^{\pi i k x}$$

$$= \alpha^{-1} e^{\pi i N x} \{ \beta_{M+1} e^{\pi i M x} - 2\beta_1 e^{\pi i x} \}$$

$$+ \alpha^{-1} e^{\pi i N x} \sum_{k=1}^{M} \beta_k e^{\pi i k x},$$

and so the spectrum of the δ_n can be expected to behave like the second sum on the right side of (2.30) and to have peaks at frequencies $x = \pi^{-1}\alpha \log p^m$ for primes p and $m \geqslant 1$.

3. Computation of the Zeros. This section describes the computations of the zeros that were used in the comparisons with eigenvalues of random Hermitian matrices described in Section 2 and documents the claim that the values that were found are accurate to within $\pm 10^{-8}$. For simplicity we will restrict the discussion here to the computations of γ_n for $n = 10^{12} + 1$ to $n = 10^{12} + 101000$. Exactly the same method was used to compute γ_n for $2001 \leqslant n \leqslant 101000$, but the error estimates in that case are much easier. (Values of γ_n for $1 \leqslant n \leqslant 2000$ were taken from the 105 decimal place values computed in [60].)

The computations described here relied on the Riemann-Siegel formula [23], [43], [68], which requires on the order of $t^{1/2}$ operation to compute $\zeta(\frac{1}{2} + it)$. A new method has recently been invented [61] which enables one to compute on the order of $T^{1/2}$ values of $\zeta(\frac{1}{2} + it)$ for $T \leqslant t \leqslant T + T^{1/2}$ in a total of $O(T^{1/2+\varepsilon})$ operations for all $\varepsilon > 0$ (i.e., $O(T^\varepsilon)$ per value), but this method is quite involved and it has yet to be implemented.

We first recall the Riemann-Siegel formula [10], [23], [28], [43], [59], [68]. Let

$$\theta(t) = \arg\left[\pi^{-it/2} \Gamma\left(\tfrac{1}{4} + it/2\right) \right],$$

and

$$Z(t) = \exp(i\theta(t))\zeta\left(\tfrac{1}{2} + it\right).$$

Then $Z(t)$ is real for real t, and any sign change of $Z(t)$ corresponds to a zero of $\zeta(s)$ on the critical line. We define

$$(3.1) \qquad \tau = (2\pi)^{-1} t, \quad m = \lfloor \tau^{1/2} \rfloor, \quad z = 2(\tau^{1/2} - m) - 1.$$

(All the computations of zeros γ_n for n near 10^{12} that were carried out had $m = 206393$.) Then, for any $k \geqslant 0$,

$$(3.2) \qquad Z(t) = 2 \sum_{n=1}^{m} n^{-1/2} \cos(t \log(n) - \theta(t))$$

$$+ (-1)^{m+1} \tau^{-1/4} \sum_{j=0}^{k} \Phi_j(z)(-1)^j \tau^{-j/2} + R_k(\tau),$$

where

(3.3) $$R_k(\tau) = O(\tau^{-(2k+3)/4}),$$

and the $\Phi_j(z)$ are certain entire functions. Gabcke [28] has derived very sharp explicit error estimates for the remainder terms $R_k(\tau)$, and the one used in the computations was

(3.4) $$|R_4(\tau)| \leqslant 0.017t^{-11/4} \quad \text{for } t \geqslant 200.$$

The $\Phi_j(z)$ were computed using their Taylor series expansion [17], [28].

Before discussing the specific implementation of the Riemann-Siegel formula that was used, we present the basic assumptions used in the error analysis. We will assume that the computer that was used (a Cray X-MP) performed as usual, i.e., that there were no intermittent faults, and the same program run on it at another time would produce the same results. We also have to make some assumptions about the correct functioning of the basic arithmetic units and the standard logarithm and cosine routines supplied by the manufacturer. It is often unsafe to trust the manufacturer's specifications, and several examples of this will be presented below.

The model of computation we will use in the error analysis of our algorithm is that developed by W. S. Brown [12]. It assumes that floating-point numbers can be represented in the signed, base b, t digit form

(3.5) $$\pm b^e \sum_{i=1}^{t} a_i b^{-i},$$

where either $a_1 = \cdots = a_t = 0$ or $1 \leqslant a_1 < b$, $0 \leqslant a_i < b$ for $2 \leqslant i \leqslant t$, and $e_{\min} \leqslant e \leqslant e_{\max}$. Brown [12] defines "correct" arithmetic for floating-point numbers in this model. If a model number is one that is representable in the form (3.5), then Brown's model can be briefly summarized by the following rules:

> If the result of applying an elementary operation ($+$, $-$, $\times$ or $/$) to two model numbers is exactly a model number, then the implementation must return that number.
> Otherwise, the implementation may return anything between the model numbers adjacent to the exact result.
> Comparison of model numbers must be exact.
> Comparison of nonmodel numbers x and y may report the result of exact comparison between any pair of elements in the smallest model intervals containing x and y.

(For more details, see [12].)

One advantage of Brown's model of computation is that it is not necessary to worry about the internal details of the floating-point arithmetic unit, but instead it is sufficient to test whether a particular machine obeys the specifications of the model. It is of course impractical to test all possible cases. N. L. Schryer [64], [65] has implemented a very extensive set of tests designed to reveal deviations from the rules of the model and has used it to test many computers. Schryer's tests are based on the fact that there tend to be patterns in the way hardward designers make mistakes, and these patterns can be detected by cleverly designed test arguments. Schryer's tests have detected most of the previously known defects as well as some new ones on a

variety of machines. The superficial description of the Cray-1 and the Cray X-MP (both of which have 64-bit words) might lead one to expect that they would satisfy Brown's model with $b = 2$, $t = 48$ for single precision and $b = 2$, $t = 96$ for double precision. However, Schryer's tests have revealed some faults. For example, multiplication of 2 by $2^{-1} + 2^{-2} + \cdots + 2^{-48}$ gives 2 on both machines. In fact, D. Winter has pointed out that it already follows from the Cray hardware manual [19] that these machines do not satisfy Brown's model for $b = 2$, $t = 48$. On the other hand, very extensive tests by Schryer's programs have shown that these machines appear to satisfy Brown's model for $b = 2$, $t = 47$ in single precision and $t = 94$ in double precision. We will assume in our analysis that this is correct. (Some defects in Cray software will be described at the end of this section.)

Using Brown's model, it is possible to bound the errors that are made in various elementary arithmetic operations. For example, if x and y are both model numbers, $2^k < x < 2^{k+1}$, $0 < y < 2^k$, and $x + y < 2^{k+1}$, where $1 \leqslant k < 40$, say, then the single-precision value z that is computed for $x + y$ will satisfy

$$|z - (x + y)| < 2^{k-46}.$$

We now describe the computation of the sum of cosines in (3.2), which is what consumes almost all of the computing time. The argument t is always maintained as a double-precision (dp) variable. Another dp variable, t_0, is also maintained, which has the property that $|t - t_0| \leqslant 3$. Three arrays d_n, q_n, u_n, $1 \leqslant n \leqslant m$, are also used; d_n is the value of $\log n$, computed and stored in dp, q_n is the value of $2n^{-1/2}$, computed in dp but stored in single precision (sp) and u_n is the value of $t_0 \log n$ reduced modulo 2π, computed in dp but again stored in sp. (On the Cray-1 and Cray X-MP, conversion from dp to sp is effected by truncation, and from sp to dp by padding with zeros.) To compute $Z(t)$ for a new value of t, a comparison of t with t_0 is made. If $|t - t_0| > 3$, t_0 is set equal to t, and the u_n are recomputed. At that point we are in the remaining case of $|t - t_0| \leqslant 3$. Here δ is defined to be the sp value of $t - t_0$, and t_1 the dp value of $t_0 + \delta$ (so that $|t_1 - t| < t_0 \cdot 2^{-90}$, say). Next, $\theta(t_1)$ is computed in dp, reduced modulo 2π, and converted to an sp variable v. Then the quantities

(3.6) $w_n = q_n \cos(\delta e_n + (u_n - v))$, $1 \leqslant n \leqslant m$,

where e_n is the sp value of d_n, are computed in sp, using the manufacturer-supplied sp cosine routine, and these values are added up in a special way to be described below.

The reason for the involved procedure described above is that for $t > 10^{10}$, say, sp arithmetic is not sufficiently accurate. On the other hand, dp arithmetic is much too slow to be used all the time, since on the Cray X-MP it is done in software and does not vectorize. Since the costly step of recomputing the u_n is done very infrequently (roughly once for every 100 evaluations of $Z(t)$, on average, in the main zero-locating run), the scheme above allows for relatively high accuracy at reasonably high speed.

The computation that is described above approximates $Z(t_1)$ rather than $Z(t)$. However, since $|t_1 - t| < t_0 \cdot 2^{-90}$, this difference is insignificant in practice, since zeros were computed only to within $\pm 5 \cdot 10^{-9}$.

We next estimate how much w_n differs from $2n^{-1/2}\cos(t_1\log n - \theta(t_1))$. The function $\theta(t_1)$ is computed entirely in dp using Stirling's formula [40]. The main term in that formula is $t_1(\log t_1 - 1 - \log(2\pi))/2$. No performance figures are supplied by the manufacturer for the dp logarithm routine. However, tests performed with 1000 random inputs $\in (0, 10^{13})$ (comparing the values obtained on the Cray X-MP with those obtained with the Macsyma symbolic manipulation system [53], which allows for variable precision) found a maximal error of $< 2 \cdot 10^{-27}$. Therefore it was assumed that all dp logarithm values (for $2 \cdot 10^{11} < t_1 < 3 \cdot 10^{11}$) are correct to within $\pm 5 \cdot 10^{-27}$. Under this assumption, the dp values of $t_1(\log t_1 - 1 - \log(2\pi))/2$ are accurate to within $\pm 4 \cdot 10^{-27}t_1$. The computation of the other terms in the expansion of $\theta(t_1)$ is easily shown to introduce an error of at most 10^{-17}. Hence after reduction modulo a dp value of 2π (correct to within $\pm 10^{-27}$), the resulting dp value of $\theta(t_1)$ is in the interval $(-0.01, 6.3)$, say, and (except for a possible additive factor of $\pm 2\pi$) is correct to within $\pm 7 \cdot 10^{-27}t_1$. The conversion to an sp value introduces an error in at most the last bit, and this error is therefore $\leqslant 2^{-44}$. Hence the sp value v of $\theta(t_1)$ differs from $\theta(t_1)$ by at most $2^{-44} + 7 \cdot 10^{-27}t_1$ (and, possibly, by an additional factor of exactly 2π). Similarly, the sp value u_n differs from $t_0 \log n$ by an integer multiple of 2π and at most $2^{-44} + 7 \cdot 10^{-27}t_1$ and again is in the range $(-0.01, 6.3)$.

Since e_n is obtained from d_n by truncation of the bit pattern, it differs from $\log n$ by at most $(2^{-47} + 10^{-24})\log n$, say. Furthermore, the sp value computed by multiplying δ and e_n in sp can differ from δe_n by at most $(2^{-47} + 2^{-93})\delta e_n$. Hence the sp value that is computed for δe_n differs from $\delta \log n$ by at most $(2^{-45} + 10^{-22})\log n$, as $|\delta| \leqslant 3$. Since $m \leqslant 210000$, we have $|\delta e_n| \leqslant 37$. Subtraction of v from u_n gives an sp value that is in the range $(-7, 7)$ but off by at most 2^{-44} from the correct value, and adding it to the sp value of δe_n yields a value that is in the interval $(-45, 45)$, but is up to $2^{-41} + 2^{-44}$ away from the correct sum of the computed values of δe_n, u_n, and $-v$. Therefore the sp value of $r_n = \delta e_n + (u_n - v)$ differs from $t_1 \log n - \theta(t_1)$ by an integer multiple of 2π and an additive factor that is

(3.7) $\leqslant 2^{-41} + 2^{-44} + (2^{-45} + 10^{-22})\log n + 2^{-43} + 1.4 \cdot 10^{-26}t_1 \leqslant 10^{-12}$

in absolute magnitude, for $t_1 \leqslant 2.7 \cdot 10^{11}$ and $n \leqslant 210000$. Hence

(3.8) $\left|\cos(r_n) - \cos(t_1\log n - \theta(t_1))\right| \leqslant 10^{-12}$.

No rigorous bounds are known for the error made by the Cray X-MP sp cosine routine. Some estimates are supplied [18, Appendix B], based on a random sample of 10^4 arguments; for example, for $x \in (-\pi, \pi)$, the maximal difference observed between $\cos(x)$ and the computed value is given as $1.01 \cdot 10^{-14}$, and the standard deviation as $2.31 \cdot 10^{-15}$. However, these figures appear to come from an old edition of the manual, whereas substantial changes have been made to the software in the meantime. A sample of 10^4 points x, drawn uniformly at random from the interval $(-\pi, \pi)$ showed that the maximal difference between values of $\cos(x)$ computed in dp and sp was only $5.34 \cdot 10^{-15}$, and the standard deviation $1.4 \cdot 10^{-15}$, which is considerably better than claimed in [18]. A similar sample of 10^6 points x, drawn uniformly at random from $(-50, 50)$ showed that the maximal difference was $\leqslant 1.35 \cdot 10^{-14}$ and the standard deviation $\leqslant 1.65 \cdot 10^{-15}$. A comparison of several

hundred values computed with the *sp* cosine routine with the same values computed in the Macsyma [53] system showed similar results. Therefore it seemed safe to assume that for $x \in (-50, 50)$, *sp* values of $\cos(x)$ are accurate to within $\pm 5 \cdot 10^{-14}$. With this assumption, the computed value of $\cos(r_n)$ differs from the correct value $\cos(t_1 \log n - \theta(t_1))$ by at most $1.05 \cdot 10^{-12}$.

Since q_n is obtained by converting the *dp* value of $2n^{-1/2}$ to *sp*, it is correct to within $\pm(2^{-47} + 2^{-90})2n^{-1/2}$, say. Therefore the *sp* value of $q_n \cos(r_n)$ that is computed differs from the desired value of $2n^{-1/2} \cos(t_1 \log n - \theta(t_1))$ by at most

$$(3.9) \qquad 2n^{-1/2}(1.05 \cdot 10^{-12} + 2^{-46} + 2^{-80}) \leqslant 2.14 \cdot 10^{-12}n^{-1/2}.$$

The sum of the term on the right side of (3.9) for $1 \leqslant n \leqslant m < 207000$ is $< 1.95 \cdot 10^{-9}$.

To add up the computed *sp* values of $q_n \cos(r_n)$, the values for $1 \leqslant n \leqslant 1280$ were converted to *dp* and added in *dp*. The n with $1280 < n \leqslant m$ were partitioned into sets B such that for each B, B had $\leqslant 120$ elements and

$$\sum_{n \in B} 2n^{-1/2} < \tfrac{1}{2}.$$

For each B, the sum S_B of the computed values of $q_n \cos(r_n)$ for $n \in B$ was evaluated in *sp*, making an error of at most 2^{-41}, converted to *dp*, and added in *dp* to the sums of the other sets B and of the initial 1280 values. There were < 3600 sets B, so the total error made in this addition is

$$< 3600 \cdot 2^{-41} + 2^{-60} < 1.7 \cdot 10^{-9}.$$

Gabcke's bound (3.4) for the remainder term in the Riemann-Siegel formula and the error terms given in [17], [28] for the Taylor series expansions of the $\Phi_j(z)$ guarantee that these sources contribute an additional error of at most $5 \cdot 10^{-11}$. We conclude therefore that the computed values of $Z(t_1)$ are correct to within $\pm 3.7 \cdot 10^{-9}$.

The actual computation of zeros was done in two stages. In the first stage, the cumbersome procedure described above for computing the sum of the w_n was replaced by a simpler one which gave approximately a 10% speedup of the entire routine for evaluating $Z(t)$. This modified routine was used to locate the zeros to a nominal accuracy (i.e., disregarding any errors in computation or from neglecting remainder terms in the Riemann-Siegel formula) of $\pm 5 \cdot 10^{-9}$. The procedure was the standard one [10], [52] of locating Gram blocks and searching for the expected number of sign changes of $Z(t)$ in them. Once these sign changes were found, zeros were computed more accurately using the Brent algorithm [9] for locating zeros of functions by a combination of linear and quadratic interpolation. On average, eight evaluations of $Z(t)$ were used for each zero.

In the second stage, the more accurate algorithm described above was used. For each value γ of a zero that was computed in the first stage, $Z(t)$ was evaluated at the two points $t = \gamma \pm 8 \cdot 10^{-9}$, the signs of the computed values were verified to be opposite, and their absolute magnitudes were checked to be greater than the upper bound $3.7 \cdot 10^{-9}$ for the error that we obtained above. In a few cases this check failed. For these zeros a more accurate version of the program, operating almost entirely in double precision, was used to verify the correctness of their locations.

The time for the first stage of the computation of the zeros was approximately 16 hours on a single-processor Cray X-MP. The same program runs only about half as fast on the Cray-1. Since the cycle time on the X-MP is only about 30% faster than on the Cray-1, most of the gain seems to be due to the larger number of memory ports on the X-MP.

The second stage, the careful verification that the computed values were as accurate as claimed, was carried out twice, once using the CFT 1.14 Fortran compiler, and once using an early unofficial version of the CFT 1.15 computer. Each run took about 5 hours. (These runs were slower than the initial zero-locating ones not only because of the less efficient method for adding up the w_n that was used, but also because the double-precision array recomputations were relatively more frequent.) The reason for undertaking two runs was that both compilers have been reported by other users to have bugs which have occasionally led to incorrect answers when operating with long vectors. The exact circumstances that lead to such wrong computations are for the most part not known very well (most seemed to be associated with long complex vectors or various optimization features that were not used in our program), and the bugs affecting the two compilers seemed different. The results of the two computations were compared, and they seemed identical to the full *sp* accuracy.

One very annoying bug in the Cray-1 and Cray X-MP software was discovered during the computations. When D is declared to be a *dp* variable, and one uses the free-format statement

$$READ *, D$$

to read a number, this number is first converted into an *sp* value, and only then converted to *dp*, so that D agrees with the value being read in only about 48 bits. The values that had been computed earlier for zeros γ_n, $10^9 \leq n \leq 10^9 + 10^5$, $10^{11} \leq n \leq 10^{11} + 10^5$, and $2 \cdot 10^{11} \leq n \leq 2 \cdot 10^{11} + 10^5$ had been converted for archival storage using a program that used the above free-format input. As a result, the values that are available for those zeros are not very accurate and were not used for most of the analyses in Section 2.

4. Statistics of Zeros. The main computation discovered 101111 zeros between Gram points g_n with $n = 10^{12} - 14$ and $n = 10^{12} + 101097$. Since the interval $[g_n, g_{n+12})$ for $n = 10^{12} - 14$ is the union of 6 Gram blocks of length 1 and 3 Gram blocks of length 2, and the interval $[g_n, g_{n+12})$ for $n = 10^{12} + 10^5 + 1085$ is the union of 8 Gram blocks of length 1 and 2 Gram blocks of length 2, we can conclude by the now standard Turing method (Theorem 3.2 of [10], for example) that the 101087 zeros that were found in the interval $[g_p, g_q)$, $p = 10^{12} - 2$, $q = 10^{12} + 10^5 + 1085$ are all the zeros of the zeta function in that range, and are indeed the zeros γ_n with $10^{12} \leq n \leq 10^{12} + 10^5 + 1086$.

The interval $[g_p, g_q)$ for $p = 10^{12} - 2$, $q = 10^{12} + 10^5 + 1$ appears to consist of 65265 Gram blocks of length 1, 10686 of length 2, 2850 of length 3, 847 of length 4, 217 of length 5, 49 of length 6, and 7 of length 7. (We say "appears" because the values of $Z(t)$ at Gram points were not checked to verify that their signs were unambiguous, except near violations of Rosser's rule.)

TABLE 7

Zeta function near selected van de Lune points.

t	$Z(t)$	Rosser rule	max δ_n
18,132,299,244.660	-133.2	no	2.95
18,139,553,794.750	142.2	no	3.08
907,663,606,940.329	229.3	no	3.34
9,065,450,718,497.579	-253.5	a	3.02
45,323,986,866,893.743	-320.7	(2,2)	3.56
67,263,231,798,214.250	441.4	b	4.11
725,177,880,629,981.915	-453.9	no	3.43

Four exceptions to Rosser's rule were found during the computation of the zeros γ_n, $10^{12} + 1 \leqslant n \leqslant 10^{12} + 10^5$. In the now standard terminology of [52], all were of length 2, and two were of type 1 ($n = 10^{12} + 15934$ and $n = 10^{12} + 93436$) and two were of type 2 ($n = 10^{12} + 13452$ and $n = 10^{12} + 17086$). Earlier computations had found 2 exceptions to Rosser's rule among zeros γ_n with $2 \cdot 10^{11} + 1 \leqslant n \leqslant 2 \cdot 10^{11} + 10^5$ (both of length 2, type 2, for $n = 2 \cdot 10^{11} + 5469$ and $n = 2 \cdot 10^{11} + 52084$), one exception among the γ_n with $10^{11} + 1 \leqslant n \leqslant 10^{11} + 10^5$ (for $n = 10^{11} + 10477$ and of length 2 and type 2), and no exceptions for $10^9 + 1 \leqslant n \leqslant 10^9 + 10^5$. Thus the frequency with which Rosser's rule is violated appears to increase very rapidly with increasing height (cf. [10], [52]).

The largest δ_n that this author is aware of is $\delta_n = 4.2626 \cdots$ for $n = 184,155,671,040$. (We have $\gamma_n \sim 5.3 \cdot 10^{10}$ here.) $Z(t)$ attains a value of approximately -214 near γ_n, and there is a violation of Rosser's rule of length 2, type 2 in the vicinity of γ_n. This zero was discovered by an application of a method for constructing values of t for which $|Z(t)|$ is large that was based on the Lovasz lattice basis reduction algorithm in a manner somewhat similar to that of [60]. Several other values of $t \lesssim 10^{11}$ with $|Z(t)|$ large were constructed that way, and a substantial fraction of them corresponded to violations of Rosser's rule, although all the violations discovered that way were of known type. It is clear that this method can be used to construct t with extremely large values of $|Z(t)|$, and the main barrier to its use is the inability to compute the corresponding values of $Z(t)$ efficiently. (It is hoped that the algorithm of [61] will be implemented in the near future, which would enable one to explore the behavior of $Z(t)$ near such special points where it is large.)

Other methods of constructing values of t with interesting behavior of $Z(t)$ are presented in [46] and [50]. Van de Lune [50] has found several values of t for which $|Z(t)|$ is very large. Thanks to D. Hejhal, the National Science Foundation, and the Minnesota Supercomputer Institute, this author was able to use a Cray-2 to study some of these points. The huge memory of the Cray-2 made it possible to utilize the algorithm described in Section 3 to compute $Z(t)$ for the most interesting of the values of t found by van de Lune. (The highest point mentioned below required about 45 million words of memory.) Rigorous error analysis was not done, and the computations spanned blocks of 20 Gram intervals centered approximately at the points found by van de Lune, so it is impossible to assert that all of the zeros in those intervals have been found and that they satisfy the RH, but it seems very likely

that they do. Table 7 presents the results of these computations. For each point t that was checked, the approximate value of $Z(t)$ is given (taken from [50], since the computations reported here have used a different grid of points at which to evaluate $Z(t)$), the type of violation of Rosser's rule that was found (if any), and the maximal δ_n that was found in that neighborhood. The value of $Z(t) \approx -453.9$ at the last point in the list gives the largest values of $|Z(t)|$ that has been found so far. An entry of "no" in the "Rosser rule" column means that no violation of Rosser's rule was found. The entry $(2, 2)$ means that a violation of length 2 and type 2 [52] was found. The entry "a" means a previously undetected type of violation of length 3, with the two Gram intervals preceding the block with zero pattern $(0, 1, 0)$ having zero pattern $(2, 2)$. The entry "b" denotes another new type of violation, this time of length 2, with Gram block with zero patterns $(0, 0)$ being followed by a block with zero pattern $(2, 1, 1, 2)$.

5. Other Methods for Computing the Zeta Function. The program described in the preceding section was designed for relatively accurate and rapid computation of the zeros of the zeta function at fairly large height. On the single-processor Cray X-MP with 2 million words of memory on which this program was run, it could only be used for zeros γ_n with n not much larger than $3 \cdot 10^{12}$, due to the limitations on memory storage. In terms of speed, on the Cray X-MP it ran at about two-thirds of the speed of the latest version [73] of the Cyber-205 program (written partially in assembly language) that was used for the verification of the RH for the first 1.5×10^9 zeros [52], since on average it required about $3.5 \cdot 10^{-3}$ seconds to evaluate a sum of the form

$$(5.1) \qquad \sum_{n=K+1}^{K+10^4} 2n^{-1/2}\cos(t\log(n) - \theta(t)),$$

whereas the data in [73] indicate that a comparable computation with the program outlined there requires about $2.5 \cdot 10^{-3}$ seconds. The program described in Section 4 has the advantages of being more portable (since it is written exclusively in Fortran) and more accurate. The error estimates in it are not completely rigorous, as the error bounds assumed for the manufacturer's cosine routine are based on statistical sampling. The main program described in [51], [52], [73] used the manufacturer's dp cosine routines only to compute a few values to medium accuracy. These values were then used inside the program to obtain sp values of the cosine by linear interpolation. Thus the [51], [52], [73] program can be regarded as somewhat more trustworthy than the one described here. (In some cases, though, where $Z(t)$ was small, a program in which all operations were carried out using the manufacturer's dp routines was used in [51], [52], and its trustworthiness is comparable to ours.)

It is possible to write programs that are much faster than that described in Section 3, are almost as accurate, are written in Fortran, and whose errors can be analyzed rigorously. Note that most of the effort in computing $Z(t)$ by the Riemann-Siegel formula (3.2) is spent in evaluating

$$\sum_{n=1}^{m} 2n^{-1/2}\cos(t\log n - \theta(t)) = \operatorname{Re} e^{-i\theta(t)} \sum_{n=1}^{m} 2n^{-1/2}e^{it\log n}.$$

Since $\theta(t)$ is easy to compute, it would suffice to find an efficient method to compute

$$(5.2) \qquad f(t) = \sum_{n=1}^{m} 2n^{-1/2}e^{it\log n}.$$

Now $m = \left\lfloor (2\pi)^{-1/2}t^{1/2} \right\rfloor$ is constant for large stretches of values of t. Suppose that m is constant over the interval $t_0 \leqslant t \leqslant t_1$, and suppose we wish to find sign changes of $Z(t)$ in this interval. Without much loss of generality we can assume that t_0 is a Gram point, $t_0 = g_k$. We compute the complex numbers

$$a_n = 2n^{-1/2}e^{it_0\log n}, \qquad 1 \leqslant n \leqslant m,$$

using some very accurate and rigorously analyzable method and store them. If $\delta = g_{k+1} - g_k$, then we also compute

$$b_{n,j} = e^{i\delta 2^{-j}\log n}, \qquad 1 \leqslant n \leqslant m, 0 \leqslant j \leqslant J,$$

(J depending on accuracy to which zeros are to be computed), and again store them. To compute $Z(t_0) = Z(g_k)$, we evaluate $a_1 + a_2 + \cdots + a_m$. To compute $Z(g_{k+1})$, we compute

$$(5.3) \qquad a_n^{(1)} = a_n \cdot b_{n,0}, \qquad 1 \leqslant n \leqslant m,$$

store them, and then evaluate $a_1^{(1)} + \cdots + a_m^{(1)}$. To compute $Z(t)$ for some t very close to g_{k+2}, we compute

$$a_n^{(2)} = a_n^{(1)} \cdot b_{n,0}, \qquad 1 \leqslant n \leqslant m,$$

store them, and then evaluate $a_1^{(2)} + \cdots + a_m^{(2)}$. (The differences $g_{l+1} - g_l$ change very slowly with l.) To compute $Z(t)$ for some t close to $(g_{k+3} + g_{k+2})/2$, we compute

$$a_n^{(3)} = a_n^{(2)} \cdot b_{n,1}, \qquad 1 \leqslant n \leqslant m,$$

store them, and evaluate $a_1^{(3)} + \cdots + a_m^{(3)}$, and so on. When we reach the point where the accumulated roundoff errors in the $a_n^{(h)}$ get too big, we reset t_0 to the value of a Gram point near t_0 and recompute the a_n. Since the recomputation of the a_n would be very infrequent in computations such as that aimed at numerically verifying the RH for a large set of consecutive zeros (where $Z(t)$ is evaluated at all Gram points), the running time of the algorithm would be dominated by the time to compute the component-by-component product $c = (c_1, \ldots, c_m) = (a_1 b_1, \ldots, a_m b_m)$ of two complex vectors $a = (a_1, \ldots, a_m)$ and $b = (b_1, \ldots, b_m)$ and then evaluate the complex sum $c_1 + \cdots + c_m$. These operations are vectorized automatically by the Cray X-MP Fortran compiler, and for $m = 10^4$ require only about 0.8×10^{-3} seconds, which would yield a program running about 4 times faster than that described in Section 3. Such a program has not been implemented owing to the large storage requirements, which would not have allowed computations with zeros around the 10^{12}th zero. However, for verifying the RH around the 10^{10}th zero, such a Fortran program ought to be practical and, since most of the computation there involves evaluating $Z(t)$ at Gram points, it would probably run about 3 times faster on the Cray X-MP than the programs described in [73] run on the Cyber-205. Some

test runs performed on the Cyber-205 by H. te Riele have shown that the ideas described above appear to lead to algorithms that are faster than those of [73] on that machine as well.

6. Distribution of Eigenvalues Computations. In the standard terminology [15], [54], [55] we let $p(k, u)$ denote the probability density of kth neighbor spacing in the GUE; i.e., the probability that the (normalized) difference between an eigenvalue of the GUE and the $(k + 1)$st smallest eigenvalue of those that are larger than it lies in the interval (a, b) is

$$(6.1) \qquad\qquad\qquad \int_a^b p(k, u)\, du.$$

(This is referred to as $p_2(k; u)$ in [15], [54], [55], where the subscript 2 refers to the GUE.)

Define λ_n and $f_n(x)$ to be the eigenvalues and eigenfunctions of the integral equation (finite Fourier transform)

$$(6.2) \qquad\qquad \lambda f(x) = \int_{-1}^1 \frac{\sin\{\pi t(x - y)/2\}}{\pi(x - y)} f(y)\, dy,$$

where $t \in \mathbf{R}^+$ is a parameter, and where the $f_n(x)$ satisfy the orthonormality condition

$$(6.3) \qquad\qquad\qquad \int_{-1}^1 f_k(x) f_m(x)\, dx = \delta_{km}.$$

The λ_n and $f_n(x)$ can be renumbered so that

$$1 \geqslant \lambda_0 \geqslant \lambda_1 \geqslant \cdots \geqslant 0.$$

Mehta and des Cloizeaux [55] have shown that

$$(6.4) \qquad p(k, t) = 4t^{-2} \prod_n (1 - \lambda_n) \sum_{0 \leqslant j_1 < \cdots < j_{k+2}} V(j_1, \ldots, j_{k+2}),$$

where

$$(6.5) \qquad V(j_1, \ldots, j_{k+2}) = 4 \prod_{i=1}^{k+2} \frac{\lambda_{j_i}}{1 - \lambda_{j_i}} \sum_{\substack{1 \leqslant \ell < m \leqslant k+2 \\ j_\ell + j_m \equiv 1 \,(\mathrm{mod}\, 2) \\ \ell < m}} f_{j_\ell}(1)^2 f_{j_m}(1)^2.$$

(Note that notations differ from reference to reference by various scaling factors.) The $p(k, t)$ were computed using Formulas (6.4) and (6.5), which are much better than some of the older methods (cf. [54], [55]) which relied on the formula

$$(6.6) \qquad\qquad\qquad p(0, t) = \frac{d^2}{dt^2} \prod_n (1 - \lambda_n).$$

For example, the values for $p(0, t)$ given in [54, Appendix A.12, Table A.1], which were apparently derived using (6.6), seem to be rather inaccurate, since for $t = 0.891$, the value given there is 0.943, whereas (6.4) and (6.5) gave the value of 0.933. The discrepancy was most pronounced near the peak of the graph of $p(0, t)$, which is where one would expect numerical differentiation routines to be least accurate.

The $f_n(x)$ are essentially the linear prolate functions. They were computed using the program of Van Buren [71], with slight modifications introduced by this author and S. P. Lloyd. In Van Buren's notation, the λ_n defined above become $\lambda_n(\pi t/2)$, and the $f_n(x)$ become $(\lambda_n(\pi t/2))^{-1/2} \psi_n(\pi t/2, x)$.

Van Buren's program uses a complicated combination of a variational procedure and expansion in terms of Legendre and spherical Bessel functions. There is no rigorous error analysis for it. However, it appears to be fairly accurate, as is shown by comparing its output to that of other calculations. Furthermore, tests such as numerical integration of

$$\int_0^\infty p(k,t)\,dt$$

to verify that we obtain values very close to 1 were all positive.

AT & T Bell Laboratories
Murray Hill, New Jersey 07974

1. R. A. BECKER & J. M. CHAMBERS, *S: An Interactive Environment for Data Analysis and Graphics*, Wadsworth, Belmont, Calif., 1984.

2. M. V. BERRY, "Semiclassical theory of spectral rigidity," *Proc. Roy. Soc. London Ser. A*, v. 400, 1985, pp. 229–251.

3. M. V. BERRY, "Riemann's zeta function: A model for quantum chaos?," in *Proc. Second Internat. Conf. on Quantum Chaos* (T. Seligman, ed.), Springer-Verlag, Berlin and New York, 1986. (To appear.)

4. O. BOHIGAS & M.-J. GIANNONI, "Chaotic motion and random matrix theories," in *Mathematical and Computational Methods in Nuclear Physics* (J. S. Dehesa, J. M. G. Gomez, and A. Polls, eds.), Lecture Notes in Phys., vol. 209, Springer-Verlag, Berlin and New York, 1984, pp. 1–99.

5. O. BOHIGAS, M. J. GIANNONI & C. SCHMIT, "Characterization of chaotic quantum spectra and universality of level fluctuation laws," *Phys. Rev. Lett.*, v. 52, 1984, pp. 1–4.

6. O. BOHIGAS, M. J. GIANNONI & C. SCHMIT, "Spectral properties of the Laplacian and random matrix theories," *J. Physique-Lettres*. (To appear.)

7. O. BOHIGAS, R. U. HAQ & A. PANDEY, "Higher-order correlations in spectra of complex systems," *Phys. Rev. Lett.*, v. 54, 1985, pp. 1645–1648.

8. E. BOMBIERI & D. HEJHAL, manuscript in preparation.

9. R. P. BRENT, *Algorithms for Minimization without Derivatives*, Prentice-Hall, Englewood Cliffs, N. J., 1973.

10. R. P. BRENT, "On the zeros of the Riemann zeta function in the critical strip," *Math. Comp.*, v. 33, 1979, pp. 1361–1372.

11. T. A. BRODY, J. FLORES, J. P. FRENCH, P. A. MELLO, A. PANDEY & S. S. M. WONG, "Random-matrix physics: spectrum and strength fluctuations," *Rev. Modern Phys.*, v. 53, 1981, pp. 385–479.

12. W. S. BROWN, "A simple but realistic model of floating-point computations," *ACM Trans. Math. Software*, v. 7, 1981, pp. 445–480.

13. J. M. CHAMBERS, W. S. CLEVELAND, B. KLEINER & P. A. TUKEY, *Graphical Methods for Data Analysis*, Wadsworth, Belmont., Calif., 1983.

14. J. DES CLOIZEAUX & M. L. MEHTA, "Some asymptotic expressions for prolate spheroidal functions and for the eigenvalues of differential and integral equations of which they are solutions," *J. Math. Phys.*, v. 13, 1972, pp. 1745–1754.

15. J. DES CLOIZEAUX & M. L. MEHTA, "Asymptotic behavior of spacing distributions for the eigenvalues of random matrices," *J. Math. Phys.*, v. 14, 1973, pp. 1648–1650.

16. J. B. CONREY, A. GHOSH, D. A. GOLDSTON, S. M. GONEK & D. R. HEATH-BROWN, "Distribution of gaps between zeros of the zeta function," *Quart. J. Math. Oxford Ser.* (2), v. 36, 1985, pp. 43–51.

17. F. D. CRARY & J. B. ROSSER, *High Precision Coefficients Related to the Zeta Function*, MRC Technical Summary Report #1344, Univ. of Wisconsin, Madison, May 1975, 171 pp.; reviewed by R. P. Brent in *Math. Comp.*, v. 31, 1977, pp. 803–804.

18. CRAY RESEARCH, INC., *Cray X-MP and Cray-1 Computer Systems; Library Reference Manual* SR-0014, Revision I, Dec. 1984.

19. CRAY RESEARCH, INC., *Cray-1 Computer Systems, S Series Mainframe Reference Manual HR*-0029, Nov. 1982.

20. D. DAVIES, "An approximate functional equation for Dirichlet *L*-functions," *Proc. Roy. Soc. Ser. A*, v. 284, 1965, pp. 224–236.

21. M. DEURING, "Asymptotische Entwicklungen der Dirichletschen *L*-Reihen," *Math. Ann.*, v. 168, 1967, pp. 1–30.

22. F. J. DYSON, "Statistical theory of the energy levels of complex systems. II," *J. Math. Phys.*, v. 3, 1962, pp. 157–165.

23. H. M. EDWARDS, *Riemann's Zeta Function*, Academic Press, New York, 1974.

24. D. FREEDMAN & P. DIACONIS, "On the histogram as a density estimator: L_2 theory," *Z. Wahrsch. Verw. Gebiete*, v. 57, 1981, pp. 453–476.

25. A. FUJII, "On the zeros of Dirichlet *L*-functions. I," *Trans. Amer. Math. Soc.*, v. 196, 1974, pp. 225–235.

26. A. FUJII, "On the uniformity of the distribution of zeros of the Riemann zeta function," *J. Reine Angew. Math.*, v. 302, 1978, pp. 167–205.

27. A. FUJII, "On the zeros of Dirichlet *L*-functions. II (with corrections to "On the zeros of Dirichlet *L*-functions. I" and the subsequent papers)," *Trans. Amer. Math. Soc.*, v. 267, 1981, pp. 33–40.

28. W. GABCKE, *Neue Herleitung und explizite Restabschätzung der Riemann-Siegel-Formel*, Ph. D. Dissertation, Göttingen, 1979.

29. P. X. GALLAGHER, "On the distribution of primes in short intervals," *Mathematika*, v. 23, 1976, pp. 4–9.

30. P. X. GALLAGHER, "Pair correlation of zeros of the zeta function," *J. Reine Angew. Math.*, v. 362, 1985, pp. 72–86.

31. P. X. GALLAGHER & J. H. MUELLER, "Primes and zeros in short intervals," *J. Reine Angew. Math.*, v. 303/304, 1978, pp. 205–220.

32. A. GHOSH, "On Riemann's zeta-function—sign changes of $S(T)$," in *Recent Progress in Analytic Number Theory*, vol. 1 (H. Halberstam and C. Hooley, eds.), Academic Press, New York, 1981, pp. 25–46.

33. A. GHOSH, "On the Riemann zeta-function—mean value theorems and the distribution of $|S(t)|$," *J. Number Theory*, v. 17, 1983, pp. 93–102.

34. D. A. GOLDSTON, "Prime numbers and the pair correlation of zeros of the zeta function," in *Topics in Analytic Number Theory* (S. W. Graham and J. D. Vaaler, eds.), Univ. of Texas Press, Austin, Texas, 1985, pp. 82–91.

35. D. A. GOLDSTON & D. R. HEATH-BROWN, "A note on the differences between consecutive primes," *Math. Ann.*, v. 266, 1984, pp. 317–320.

36. D. A. GOLDSTON & H. L. MONTGOMERY, "Pair correlation of zeros and primes in short intervals," *Proc. 1984 Stillwater Conf. on Analytic Number Theory and Diophantine Problems*, Birkhäuser-Verlag. (To appear.)

37. S. M. GONEK, "A formula of Landau and mean values of $\zeta(s)$," in *Topics in Analytic Number Theory* (S. W. Graham and J. D. Vaaler, eds.), Univ. of Texas Press, Austin, Texas, 1985, pp. 92–97.

38. A. P. GUINAND, "A summation formula in the theory of prime numbers," *Proc. London Math. Soc.* (2), v. 50, 1948, pp. 107–119.

39. M. C. GUTZWILLER, "Stochastic behavior in quantum scattering," *Phys. D*, v. 7, 1983, pp. 341–355.

40. HANDBOOK OF MATHEMATICAL FUNCTIONS (M. Abramowitz and I. A. Stegun, eds.), National Bureau of Standards, Washington, D.C., 9th printing, 1970.

41. R. U. HAQ, A. PANDEY & O. BOHIGAS, "Fluctuation properties of nuclear energy levels: Do theory and experiment agree?," *Phys. Rev. Lett.*, v. 48, 1982, pp. 1086–1089.

42. D. R. HEATH-BROWN, "Gaps between primes and the pair correlation of zeros of the zeta-function," *Acta Arith.*, v. 41, 1982, pp. 85–99.

43. A. IVIC, *The Riemann Zeta-function*, Wiley, New York, 1985.

44. D. JOYNER, "Distribution theorems for *L*-functions." (To be published.)

45. D. JOYNER, "On the Dyson-Montgomery hypothesis." (To be published.)

46. E. KARKOSCHKA & P. WERNER, "Einige Ausnahmen zur Rosserschen Regel in der Theorie der Riemannschen Zetafunktion," *Computing*, v. 27, 1981, pp. 57–69.

47. M. G. KENDALL & A. STUART, *The Advanced Theory of Statistics*, 3rd ed., Hafner, New York, 1973.

48. E. LANDAU, "Über die Nullstellen der Zetafunktion," *Math. Ann.*, v. 71, 1911, pp. 548–564.

49. R. S. LEHMAN, "On the distribution of zeros of the Riemann zeta function," *Proc. London Math. Soc.*, v. 20, 1970, pp. 303–320.

50. J. VAN DE LUNE, *Some Observations Concerning the Zero-Curves of the Real and Imaginary Parts of Riemann's Zeta Function*, Report ZW 201/83, Mathematical Center, Amsterdam, December 1983.

51. J. VAN DE LUNE, H. J. J. TE RIELE & D. T. WINTER, *Rigorous High Speed Separation of Zeros of Riemann's Zeta Function*, Report NW 113/81, Mathematical Center, Amsterdam, 1981.

52. J. VAN DE LUNE, H. J. J. TE RIELE & D. T. WINTER, "On the zeros of the Riemann zeta function in the critical strip. IV," *Math. Comp.*, v. 46, 1986, pp. 667–681.

53. MATHLAB GROUP, *MACSYMA Reference Manual*, MIT Laboratory for Computer Science, 1977.

54. M. L. MEHTA, *Random Matrices*, Academic Press, New York, 1967.

55. M. L. MEHTA & J. DES CLOIZEAUX, "The probabilities for several consecutive eigenvalues of a random matrix," *Indian J. Pure Appl. Math.*, v. 3, 1972, pp. 329–351.

56. H. L. MONTGOMERY, *The Pair Correlation of Zeros of the Zeta Function*, Proc. Sympos. Pure Math., vol. 24, Amer. Math. Soc., Providence, R. I., 1973, pp. 181–193.

57. H. L. MONTGOMERY, "Distribution of zeros of the Riemann zeta function," *Proc. Internat. Congress Math. Vancouver*, 1974, pp. 379–381.

58. H. L. MONTGOMERY, "Extreme values of the Riemann zeta function," *Comment. Math. Helv.*, v. 52, 1977, pp. 511–518.

59. A. M. ODLYZKO, "Distribution of zeros of the Riemann zeta function: Conjectures and computations." (Manuscript in preparation.)

60. A. M. ODLYZKO & H. J. J. TE RIELE, "Disproof of the Mertens conjecture," *J. Reine Angew. Math.*, v. 357, 1985, pp. 138–160.

61. A. M. ODLYZKO & A. SCHÖNHAGE, "Fast algorithms for multiple evaluations of the Riemann zeta function." (To be published.)

62. A. E. OZLUK, *Pair Correlation of Zeros of Dirichlet L-functions*, Ph. D. Dissertation, Univ. of Michigan, Ann Arbor, Mich., 1982.

63. C. E. PORTER, ed., *Statistical Theories of Spectra: Fluctuations*, Academic Press, New York, 1965.

64. N. L. SCHRYER, *A Test of a Computer's Floating-Point Arithmetic Unit*, AT & T Bell Laboratories Computing Science Technical Report #89, 1981.

65. N. L. SCHRYER, manuscript in preparation.

66. A. SELBERG, "Contributions to the theory of the Riemann zeta-function," *Arch. Math. Naturvid. B*, v. 48, 1946, pp. 89–155.

67. C. L. SIEGEL, "Über Riemanns Nachlass zur analytischen Zahlentheorie," *Quellen und Studien zur Geschichte der Math. Astr. Phys.*, v. 2, 1932, pp. 45–80; reprinted in C. L. Siegel, *Gesammelte Abhandlungen*, vol. 1, Springer-Verlag, Berlin and New York, 1966, pp. 275–310.

68. E. C. TITCHMARSH, *The Theory of the Riemann Zeta-function*, Oxford Univ. Press, Oxford, 1951.

69. K.-M. TSANG, *The Distribution of the Values of the Riemann Zeta-function*, Ph. D. Dissertation, Princeton, 1984.

70. A. M. TURING, "A method for the calculation of the zeta-function," *Proc. London Math. Soc.* (2), v. 48, 1943, pp. 180–197.

71. A. L. VAN BUREN, *A Fortran Computer Program for Calculating the Linear Prolate Functions*, Report 7994, Naval Research Laboratory, Washington, May 1976.

72. A. WEIL, "Sur les "formules explicites" de la théorie des nombres premiers," *Comm. Sém. Math. Univ. Lund*, tome supplémentaire, 1952, pp. 252–265.

73. D. WINTER & H. TE RIELE, Optimization of a program for the verification of the Riemann hypothesis, *Supercomputer*, v. 5, 1985, pp. 29–32.